OM nucleic - nucleic search, using sw model

GenCore version 5.1.4_p5_4578
Copyright (c) 1993 - 2003 Compugen Ltd.

Run on: March 29, 2003, 08:02:28 ; Search time 272.803 Seconds
(without alignments)
8255.046 Million cell updates/sec

Title: US-09-691-219-3_COPY_1_1000
Perfect score: 1000
Sequence: 1 cctgccaccatgcctggcta.........caagaggcctgcacctctga 1000

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 2185239 seqs, 112599159 residues

Total number of hits satisfying chosen parameters: 4370478

Post-processing: Minimum Match 08
Maximum Match 100%
Listing first 45 summaries

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Database : N_Geneseq_1010002:*
```
 1: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1980.DAT:*
 2: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1981.DAT:*
 3: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1982.DAT:*
 4: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1983.DAT:*
 5: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1984.DAT:*
 6: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1985.DAT:*
 7: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1986.DAT:*
 8: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1987.DAT:*
 9: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1988.DAT:*
10: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1989.DAT:*
11: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1990.DAT:*
12: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1991.DAT:*
13: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1992.DAT:*
14: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1993.DAT:*
15: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1994.DAT:*
16: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1995.DAT:*
17: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1996.DAT:*
18: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1997.DAT:*
19: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1998.DAT:*
20: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA2000.DAT:*
21: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA2001A.DAT:*
22: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA2001B.DAT:*
23: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA2002.DAT:*
```

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	1000	100.0	11101	22	ABN83947	Human transporter
2	226	22.6	10702	22	AAK64975	Human immune/haema
3	217.6	21.8	59065	24	ABK42416	Human serine/threo
4	217.6	21.8	128978	24	ABK83459	Human cDNA differe
5	211	21.1	13216	22	AAL05122	Human reproductive
6	211	21.1	13216	23	ABL98014	Human testicular a
7	209.8	21.0	169739	24	ABQ08186	Human osteoblast d
8	209.4	20.9	29163	22	AAL05121	Human reproductive
9	209.4	20.9	29163	23	ABL98013	Human testicular a
c 10	207.8	20.8	122888	24	ABK83569	Human cDNA differe
c 11	206.2	20.6	946	22	AAK64975	Human immune/haema
c 12	205.8	20.6	25715	22	AAS33462	DNA encoding human
13	205.8	20.6	48908	22	AAK82338	Human nervous syst
14	202	20.2	32192	22	ABA15328	Human immune/haema
15	202	20.2	32192	22	AAL05257	Human reproductive
16	202	20.2	32192	22	AAL05532	Human reproductive
17	202	20.2	32192	23	ABL05631	Human reproductive
c 18	202	20.2	32192	23	ABL98141	Human testicular a
19	202	20.2	32192	23	ABK42529	Human testicular a
c 20	201.8	20.2	11538	22	AAS32726	Genomic sequence #
21	201.8	20.2	11538	22	AAS36098	Human genomic DNA
22	201.4	20.1	7703	22	AAK90904	Human cardiovascul
23	201.4	20.1	7703	22	AAX90945	Human digestive sy
24	201.4	20.1	7703	22	AAS31939	Human digestive sy
25	201.4	20.1	7703	22	AAS31980	Human liver associ
26	201.4	20.1	7703	22	ABN90294	Human liver associ
27	201.4	20.1	7703	24	ABN90335	Human liver antige
28	201.4	20.1	4282	23	AAL41898	Human liver antige
c 29	199.6	20.0	7703	24	ABN90335	Human GA733-2 gene
30	199.2	19.9	65608	24	ABL62910	Breast cancer rela
31	199.2	19.9	65608	24	ABL64414	Stomach cancer rel
32	199.2	19.9	65608	24	ABL67698	Oesophagus cancer
33	198.6	19.9	65608	22	AAK70995	Human genomic DNA
34	197.8	19.8	19199	22	AAS32785	Human immune/haema
35	197.8	19.8	11853	22	AAK77832	Human immune/haema
36	197.4	19.7	25895	22	AAK77833	Human immune/haema
37	197.4	19.7	44354	22	AAK77836	Human immune/haema
38	197.4	19.7	44354	22	AAK77837	Human immune/haema
39	197.4	19.7	12394	22	AAD14749	Human glycogen syn
40	197.2	19.7	14737	22	AAK86397	Human immune/haema
41	196.8	19.7	26191	22	AAK86398	Human immune/haema
42	196.6	19.7	109906	24	ABK94411	DNA encoding endot
43	196.4	19.6	32177	22	ABA18493	Human nervous syst
44	196.4	19.6	32177	22	AAS29828	Human cytoskeletal
44	196.4	19.6	32177	22	AAL05374	Human reproductive
45	195.6	19.6	22081	22	AAF97862	Human neuroblastom

ALIGNMENTS

RESULT 1
ABN83947

```
ID   ABN83947 standard; DNA; 11101 BP.
XX
AC   ABN83947;
XX
DT   02-OCT-2002 (first entry)
XX
DE   Human transporter protein encoding genomic DNA.
XX
KW   Human; transporter protein; zinc transporter; pharmacogenomic analysis;
KW   diagnosis; drug screening; gene therapy; kidney; testis; heart; placenta;
KW   small intestine; liver; chromosome 1; single nucleotide polymorphism;
KW   SNP; gene; ds.
XX
OS   Homo sapiens.
XX
FH   Key       Location/Qualifiers
FH
FT   CDS       2284..8840
FT             /*tag= a
FT   exon      2284..2333
FT             /*tag= b
FT             /product= "transporter protein"
FT             /note= "contains 7 introns"
FT   intron    2334..2962
FT             /*tag= c
FT             /number= 1
FT   exon      2963..3183
FT             /*tag=
FT             /number= 2
FT   intron    3184..3735
```

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

```
FT                   /*tag= e
FT   exon            /number= 2
FT                   3736..3882
FT   exon            /*tag= f
FT                   /number= 3
FT                   3883..4629
FT   intron          /*tag= g
FT                   /number= 3
FT                   4630..4783
FT   exon            /*tag= h
FT                   /number= 4
FT                   4784..5472
FT   intron          /*tag= j
FT                   /number= 4
FT                   5473..5632
FT   exon            /*tag= j
FT                   /number= 5
FT                   5633..6375
FT   intron          /*tag= k
FT                   /number= 5
FT                   6376..6481
FT   exon            /*tag= l
FT                   /number= 6
FT                   6482..8088
FT   intron          /*tag= m
FT                   /number= 6
FT                   8089..8223
FT   exon            /*tag= n
FT                   /number= 7
FT                   8224..8694
FT   intron          /*tag= o
FT                   /number= 7
FT                   8695..8838
FT   exon            /*tag= p
FT                   /number= 8
XX
PN   WO200224910-A2.
XX
PD   28-MAR-2002.
XX
PF   20-SEP-2001; 2001WO-US29218.
XX
PR   20-SEP-2000; 2000US-234160P.
PR   19-OCT-2000; 2000US-0691219.
XX
PA   (PEKE ) PE CORP NY.
XX
PI   Wei M, Ketchum KA, Di Francesco V, Beasley EM;
XX
DR   WPI; 2002-404954/43.
DR   P-PSDB; ABB83084.
XX
PT   Novel human transporter proteins, related to zinc transporter
PT   subfamily, useful as model for developing human therapeutic targets and
PT   serves as target for human therapeutics -
XX
PS   Claim 4(b); Fig 3; 75pp; English.
XX
CC   The invention relates to an isolated human transporter protein that is
CC   related to the zinc transporter subfamily. Polynucleotides and
CC   polypeptides of the invention are useful for treating a disease or
CC   condition mediated by human transporter protein. The proteins also
CC   provide a target for diagnosing a disease or predisposition to disease
CC   mediated by the peptide, and in pharmacogenomic analysis. The peptides
CC   are also useful for treating a disorders characterised by absence of,
CC   inappropriate or unwanted expression of the protein. The nucleic acids
CC   are also useful in drug screening assays and as a target for treatment
CC   by the compounds identified through drug screening. The invention also
CC   provides vectors for gene therapy in patients with aberrant expression
CC   of the gene encoding the transporter protein. The gene of the invention
CC   has been found to be expressed in humans in the kidney, testis, heart,
CC   placenta, small intestine and liver. The gene has been localised to
CC   human chromosome 1. The current sequence represents the human transporter
```

```
CC   protein encoding genomic DNA. This sequence contains 6 single nucleotide polymorphisms (SNP's),
CC   Note: This sequence contains 6 single nucleotide polymorphisms (SNP's),
CC   but information given in the specification is insufficient to determine
CC   their locations within this sequence (see ABN83948-ABN83953 for specific
CC   regions).
XX
SQ   Sequence 11101 BP; 2403 A; 3041 C; 3200 G; 2448 T; 9 other;

Query Match            100.0%;  Score 1000;  DB 24;  Length 11101;
Best Local Similarity  100.0%;  Pred. No. 1.1e-240;
Matches 1000;  Conservative   0;  Mismatches   0;  Indels   0;  Gaps   0;

Qy         1 CCTGCCACCATGCCTGGCTAATTTTCTTATTTTTAGTAGAGACGAGGTTTGGCCATGTTG 60
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         1 CCTGCCACCATGCCTGGCTAATTTTCTTATTTTTAGTAGAGACGAGGTTTGGCCATGTTG 60

Qy        61 ACCAGGCTGGTCTCGAACTCCTGGTCTTGGGTGGGCCTGCTGCCTCAGCCTCCCAAAGTG 120
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        61 ACCAGGCTGGTCTCGAACTCCTGGTCTTGGGTGGGCCTGCTGCCTCAGCCTCCCAAAGTG 120

Qy       121 CTGGGATTATAGGCGTGAGCCCCCGCCCCGCACCCAGCCCAGCCCAAACATTTTTAAAGTAG 180
             |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       121 CTGGGATTATAGGCGTGAGCCCCCGCCCCGCACCCAGCCCAGCCCAAACATTTTTAAAGTAG 180

Qy       181 AGGGAATAGTTATAGTGTACCCCCATTTACCCATCACTCAGTTTCAACAGCTGGTTGACTA 240
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       181 AGGGAATAGTTATAGTGTACCCCCATTTACCCATCACTCAGTTTCAACAGCTGGTTGACATA 240

Qy       241 TTTATTTCTTCTTATACCAGTACCGTACTCTCCCCACTGGGATTTATTTTTTAGGCAAACC 300
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       241 TTTATTTCTTCTTATACCAGTACCGTACTCTCCCCACTGGGATTTATTTTTTAGGCAAACCC 300

Qy       301 AGATGACACATTTTATCCCTAAATACTTTAGATAAAGGTGTTTCTTTTGAAAAAATTCATAACC 360
             |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       301 AGATGACACATTTTATCCCTAAATACTTTAGATAAAGGTGTTTCTTTTGAAAAAATCATAACC 360

Qy       361 TCAGGACCCAGCCTGGCCAACATGGTGAAACCCCGTCTCTGTACTAAAAATACAAAAATTAGC 420
             |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       361 TCAGGACCCAGCCTGGCCAACATGGTGAAACCCCGTCTCTGTACTAAAAATACAAAAATTAGC 420

Qy       421 TTGGCATGGTGGTGCGTGGGCCACCTGTAATCCCAGCTACTCAGGAGGCTGAGGCAGGAATCA 480
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       421 TTGGCATGGTGGTGCGTGGGCACCTGTAATCCCAGCTACTCAGGAGGCTGAGGCAGGAATCA 480

Qy       481 CTTGAATCCGGGAAGCAGGAGATTGCAGTGAGCCGAGATCGAGCCTGGGCGACAG 540
             |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       481 CTTGAATCCGGGAAGCAGGAGATTGCAGTGAGCCGAGATCGAGCCTGGGCGACAG 540

Qy       541 AGACAGAAAATGAAACTCTGTCTCTCAAAAACAAAACAAAAAAAACCACTATACATAAAAA 600
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       541 AGACAGAAAATGAAACTCTGTCTCTCAAAAACAAAACAAAAAAAACCACTATACATAAAAA 600

Qy       601 TGAACAAATGATGCCACAATAGCACCAGAGAATTTTATAAATACAGATTCCAGCCCTGC 660
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       601 TGAACAAATGATGCCACAATAGCACCAGAGAATTTTATAAATACAGATTCCAGCCCTGC 660

Qy       661 CCCAGACCTACTGAATCCTGGAAATATTCAGGCTTCCACACCCAGAGAATTCTGGTTCGGTT 720
             |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       661 CCCAGACCTACTGAATCCTGGAAATATTCAGGCTTCCACACCCAGAGAATTCTGGTTCGGTT 720

Qy       721 GGTCTGATCGCAGGGACCTGTAACCTGCGTTGTAACACCTTCTCAGGTAATGCTGAGCCT 780
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       721 GGTCTGATCGCAGGGACCTGTAACCTGCGTTGTAACACCTTCTCAGGTAATGCTGAGCCT 780

Qy       781 GCTGGTGCTCAGAGTAGACAGACCTGGAGAAAACCAGGTGTCTGTGAGGTTTTTCCAGAGA 840
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       781 GCTGGTGCTCAGAGTAGACAGACCTGGAGAAAACCAGGTGTCTGTGAGGTTTTTCCAGAGA 840

Qy       841 AAACCAGAGCTCCAGAGAGGACCCACTCAGTGAGGACCCAAGCAGAGCAGAGGGGTGCAC 900
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       841 AAACCAGAGTCCAGAGAAGGACCCACTCAGTGAGGACCCAAGCAGAGCAGAGGGGTGCAC 900

Qy       901 CTCACATCCTCACTCCTGGCACCCGTCTCCTACAAGATGAGAGACTGAAAGAGAGCCCTT 960
```

```
Db        901 CTCACATCCTCACTCCTGCCACCCGTCTCCTACACGATGAGAGAGACTGAAAGAGCCCCTT 960
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy        961 CCTGTCCCCAGTGTGTGGGCAAGAGGAGGCCTGCACCTCTGA 1000
              |||||||||||||||||||||||||||| |||||||||||||
Db        961 CCTGTCCCCAGTGTGTGGGCAAGAGGGCCTGCACCTCTGA 1000

RESULT 2
AAK69757/c
ID   AAK69757 standard; DNA; 10702 BP.
XX
AC   AAK69757;
XX
DT   06-NOV-2001 (first entry)
XX
DE   Human immune/haematopoietic antigen genomic sequence SEQ ID NO:24569.
XX
KW   Human; immune; haematopoietic; immune/haematopoietic antigen; cancer;
KW   cytostatic; gene therapy; vaccine; metastasis; ds.
XX
OS   Homo sapiens.
XX
PN   WO20015?182-A2.
XX
PD   09-AUG-2001.
XX
PF   17-JAN-2001; 2001WO-US01354.
XX
PR   31-JAN-2000; 2000US-0179065.          PR   08-SEP-2000; 2000US-0231242.
PR   04-FEB-2000; 2000US-0180628.          PR   08-SEP-2000; 2000US-0231243.
PR   24-FEB-2000; 2000US-0184664.          PR   08-SEP-2000; 2000US-0231244.
PR   02-MAR-2000; 2000US-0186350.          PR   08-SEP-2000; 2000US-0231413.
PR   02-MAR-2000; 2000US-0189874.          PR   08-SEP-2000; 2000US-0231414.
PR   16-MAR-2000; 2000US-0190076.          PR   08-SEP-2000; 2000US-0232080.
PR   17-MAR-2000; 2000US-0198123.          PR   08-SEP-2000; 2000US-0232081.
PR   18-APR-2000; 2000US-0198123.          PR   08-SEP-2000; 2000US-0231968.
PR   19-MAY-2000; 2000US-0205515.          PR   12-SEP-2000; 2000US-0232397.
PR   07-JUN-2000; 2000US-0209467.          PR   14-SEP-2000; 2000US-0232398.
PR   28-JUN-2000; 2000US-0214886.          PR   14-SEP-2000; 2000US-0232399.
PR   30-JUN-2000; 2000US-0215135.          PR   14-SEP-2000; 2000US-0232400.
PR   07-JUL-2000; 2000US-0216647.          PR   14-SEP-2000; 2000US-0232401.
PR   07-JUL-2000; 2000US-0216880.          PR   14-SEP-2000; 2000US-0233063.
PR   11-JUL-2000; 2000US-0217487.          PR   14-SEP-2000; 2000US-0233064.
PR   11-JUL-2000; 2000US-0217496.          PR   14-SEP-2000; 2000US-0233065.
PR   14-JUL-2000; 2000US-0218290.          PR   21-SEP-2000; 2000US-0234223.
PR   26-JUL-2000; 2000US-0220963.          PR   21-SEP-2000; 2000US-0234274.
PR   26-JUL-2000; 2000US-0220964.          PR   25-SEP-2000; 2000US-0234997.
PR   14-AUG-2000; 2000US-0224518.          PR   25-SEP-2000; 2000US-0234998.
PR   14-AUG-2000; 2000US-0224519.          PR   26-SEP-2000; 2000US-0235484.
PR   14-AUG-2000; 2000US-0225213.          PR   27-SEP-2000; 2000US-0235834.
PR   14-AUG-2000; 2000US-0225214.          PR   27-SEP-2000; 2000US-0235836.
PR   14-AUG-2000; 2000US-0225266.          PR   29-SEP-2000; 2000US-0236327.
PR   14-AUG-2000; 2000US-0225267.          PR   29-SEP-2000; 2000US-0236367.
PR   14-AUG-2000; 2000US-0225268.          PR   29-SEP-2000; 2000US-0236368.
PR   14-AUG-2000; 2000US-0225270.          PR   29-SEP-2000; 2000US-0236369.
PR   14-AUG-2000; 2000US-0225447.          PR   29-SEP-2000; 2000US-0236370.
PR   14-AUG-2000; 2000US-0225757.          PR   02-OCT-2000; 2000US-0236802.
PR   14-AUG-2000; 2000US-0225758.          PR   02-OCT-2000; 2000US-0237037.
PR   14-AUG-2000; 2000US-0225759.          PR   02-OCT-2000; 2000US-0237038.
PR   18-AUG-2000; 2000US-0226279.          PR   02-OCT-2000; 2000US-0237039.
PR   22-AUG-2000; 2000US-0226681.          PR   02-OCT-2000; 2000US-0237040.
PR   22-AUG-2000; 2000US-0226868.          PR   13-OCT-2000; 2000US-0239935.
PR   22-AUG-2000; 2000US-0227182.          PR   20-OCT-2000; 2000US-0239937.
PR   23-AUG-2000; 2000US-0227009.          PR   20-OCT-2000; 2000US-0240960.
PR   30-AUG-2000; 2000US-0228924.          PR   20-OCT-2000; 2000US-0241221.
PR   01-SEP-2000; 2000US-0229287.          PR   20-OCT-2000; 2000US-0241785.
PR   01-SEP-2000; 2000US-0229343.          PR   20-OCT-2000; 2000US-0241786.
PR   01-SEP-2000; 2000US-0229344.          PR   20-OCT-2000; 2000US-0241787.
PR   01-SEP-2000; 2000US-0229345.          PR   20-OCT-2000; 2000US-0241808.
PR   05-SEP-2000; 2000US-0229509.          PR   20-OCT-2000; 2000US-0241809.
PR   05-SEP-2000; 2000US-0229513.          PR   01-OCT-2000; 2000US-0241826.
PR   06-SEP-2000; 2000US-0230437.          PR   20-OCT-2000; 2000US-0244617.
PR   06-SEP-2000; 2000US-0230438.          PR   08-NOV-2000; 2000US-0246474.
                                           PR   08-NOV-2000; 2000US-0246475.
                                           PR   08-NOV-2000; 2000US-0246476.
                                           PR   08-NOV-2000; 2000US-0246477.
                                           PR   08-NOV-2000; 2000US-0246478.
                                           PR   08-NOV-2000; 2000US-0246523.
                                           PR   08-NOV-2000; 2000US-0246524.
                                           PR   08-NOV-2000; 2000US-0246525.
                                           PR   08-NOV-2000; 2000US-0246526.
                                           PR   08-NOV-2000; 2000US-0246527.
                                           PR   08-NOV-2000; 2000US-0246528.
                                           PR   08-NOV-2000; 2000US-0246532.
                                           PR   08-NOV-2000; 2000US-0246609.
                                           PR   08-NOV-2000; 2000US-0246610.
                                           PR   08-NOV-2000; 2000US-0246611.
                                           PR   08-NOV-2000; 2000US-0246613.
                                           PR   17-NOV-2000; 2000US-0249207.
                                           PR   17-NOV-2000; 2000US-0249208.
                                           PR   17-NOV-2000; 2000US-0249209.
                                           PR   17-NOV-2000; 2000US-0249210.
                                           PR   17-NOV-2000; 2000US-0249211.
                                           PR   17-NOV-2000; 2000US-0249212.
                                           PR   17-NOV-2000; 2000US-0249213.
                                           PR   17-NOV-2000; 2000US-0249214.
                                           PR   17-NOV-2000; 2000US-0249215.
                                           PR   17-NOV-2000; 2000US-0249216.
                                           PR   17-NOV-2000; 2000US-0249217.
                                           PR   17-NOV-2000; 2000US-0249218.
                                           PR   17-NOV-2000; 2000US-0249244.
```

```
PR   17-NOV-2000;  2000US-0249245.
PR   17-NOV-2000;  2000US-0249264.
PR   17-NOV-2000;  2000US-0249265.
PR   17-NOV-2000;  2000US-0249297.
PR   17-NOV-2000;  2000US-0249299.
PR   17-NOV-2000;  2000US-0249300.
PR   01-DEC-2000;  2000US-0250160.
PR   01-DEC-2000;  2000US-0250391.
PR   05-DEC-2000;  2000US-0251030.
PR   05-DEC-2000;  2000US-0251988.
PR   05-DEC-2000;  2000US-0256719.
PR   05-DEC-2000;  2000US-0257119.
PR   06-DEC-2000;  2000US-0251479.
PR   08-DEC-2000;  2000US-0251856.
PR   08-DEC-2000;  2000US-0251868.
PR   08-DEC-2000;  2000US-0251869.
PR   08-DEC-2000;  2000US-0251989.
PR   08-DEC-2000;  2000US-0251990.
PR   11-DEC-2000;  2000US-0254097.
PR   05-JAN-2001;  2001US-0259678.
XX
PA   (HUMA-) HUMAN GENOME SCI INC.
XX
PI   Rosen CA,  Barash SC,  Ruben SM;
XX
DR   WPI; 2001-483426/52.
XX
PS   Disclosure; SEQ ID NO 24569; 3071pp + Sequence Listing; English.
XX
PT   Nucleic acids encoding human immune/hematopoietic antigen polypeptides,
PT   useful for preventing, diagnosing and/or treating cancers and
PT   metastasis -
XX
CC   AAK54951 to AAK64702 encode the human immune/haematopoietic antigen (I)
CC   amino acid sequences given in AAM82170 to AAM91921. (I) have cytostatic
CC   activity, and can be used in gene therapy and vaccine production. (I)
CC   proteins and polynucleotides may be used in the prevention, diagnosis and
CC   treatment of diseases associated with inappropriate (I) expression. For
CC   example, they may be used to treat disorders associated with decreased
CC   expression by rectifying mutations or deletions in a patient's genome
CC   that affect the activity of (I) by expressing inactive proteins or to
CC   supplement the patients own production of (I). Additionally, (I)
CC   polynucleotides may be used to produce the secreted (I), by inserting
CC   the nucleic acids into a host cell and culturing the cell to express the
CC   protein. (I) proteins and polynucleotides may be used to prevent,
CC   diagnose and treat immune/haematopoietic-related diseases, especially
CC   cancers and cancer metastases of haematopoietic-derived cells. AAK64703
CC   to AAK87694 represent human immune/haematopoietic antigen genomic
CC   sequences from the present invention. AAK54942 to AAK54950 and AAM82169
CC   represent sequences used in the exemplification of the present invention.
XX
SQ   Sequence 10702 BP; 2680 A; 2543 C; 2609 G; 2870 T; 0 other;

Query Match             22.6%;  Score 226;  DB 22;  Length 10702;
Best Local Similarity   64.4%;
Matches  390;  Conservative    0;  Mismatches 205;  Indels  11;  Gaps   3;

Qy      4  GCCACCATGCCTGGCTAATTTTTCTTATTTTTAGTAGAGACGAGGTTTTGCCATGTTGACC  63
            | |||||| |  |||||||||||| | ||||||||||||||||||||| ||||||||||||
Db    910  GTCACCAAGCCCAGCCTAGTTTTTGTATTTTTAGCAAAGACGGGGTTTCACCATGTTGCC  851

Qy     64  AGGCTGGTCTCGAAACYCTTGACCTCAGGTGATCCAGCCTCCCAAAGTGCTG  123
            ||||||||||||  ||| |||||||||||||||||||||||||||||||||||
Db    850  AGGCTGGTCTTGAACTCCTGACCTCAGGTGATCCAGCCTCCCAAAGTGCTG  791

Qy    124  GGATTATAGGGGTGAGCCGCGCGCCCACCCAGCCAACATTTTTTAAATACTGAAAAGTAGAGG  183
            |||||||| |||||| ||| ||||||||||||| ||||||| |||| || ||||||||| ||||
Db    790  GGATTACAGGCGTTGAGCCACCGCGCCCGGCCTGCTCTGCATATTTTATGTCTGTATTCT  731

Qy    184  GAATAGTTATAGTGTACCCCATTTACCCATCTCAGTTTCAACAGCTGGTGACATATTT  243
            |||| |    |  |     |  ||||||||||||||||||||||||||||||||| |||
Db    730  GTGCATATTCTTCTCAGCAGATAATAATAAAAACTTCACAGAGGGGCTCCAGACACTGGGGCTT  671

Qy    244  ATTTCTTCTATACCAGTACCGTACTCTCCCCACTGGGATTATTTATTTAAGGCAAAAACC--CA  301
```

```
Db    670  AATAGTCTTAAAAAATATATAAATATATATATATATTATTTTTTTTTCAGAAATATATTTCCAGTA  611
            | ||| ||||||||||||||||||||||||||||| |||| || |||||||||||||  || |
Db    302  GATGACACATTTTATCCCTAAATACTTTAGATAAAGGTGTTCTTTGAAAAAAATCATAACCT  361
            || || | |  |||||  |||  || || |||||| ||  ||| ||| |||  ||| || |
Db    610  GAGAACTAGAAATCATAAAGAAAAAAATTAGCGGACGCGGATCACCTGAGGTCAGGAGTT  551
            ||||| ||||  |||  || |||||||||||| | ||||||| |||| |||||||||| || |
Qy    362  CAGGACCAGCCCTGCCCACATGGTGAAACCCTGTCTGTACTAAAAATACAAAAATTAGCT  421
            || |||| |||| |||| | |||||||||| |||| | |  ||||| ||||||||| |||||
Db    550  CAAGACCAGCCCTGCCGACATGGCGAAACCCCATCTCTACTAAAAATACAAAAATTAGCC  491
            || |||| |||||||  || ||| | |||||  |||| | |||||||||||||||| ||||
Qy    422  TGGGCATGGTGGCGTGGGCACCTGTAATCCCAGCTACTCAGGAAGGCTGAGGCAGGAGAATCAC  481
            |||||| | ||| |||| | |||| || |||| ||| | |||||| ||||||||| |||||||
Db    490  GGGCTTGGTGGTGGGCGGCTGTGTTAATCCCAGCTACTCAGGGAGGCTGAGGCAGGAGAATCAC  431
            | |||  || ||| | | | ||| || |||| ||||||||||  |||||||||||||||||||
Qy    482  TTGAATCCGGGAACCAGAGATTGCAGTGAGCTGAGA---TTGCAGTCGAGCCTGGGCGAC  538
            || ||||| | | | |||| ||||| ||||| |||     |||| | |||| |||| |||
Db    490  GGGCTTGGTGGGGCGGCTGTGTTAATCCCAGCTACTCAGGGAGGCTGAGGCAGGAGAATCAC  431
            ...
```

```
RESULT 3
ABL42416
ID   ABL42416  standard; DNA; 59065 BP.
XX
AC   ABL42416;
XX
DT   11-SEP-2002  (first entry)
XX
DE   Human serine/threonine kinase DNA.
XX
KW   Human;  serine/threonine kinase; LIM domain kinase;
KW   LIMK2b homologue; drug discovery; SNP;
KW   single nucleotide polymorphism; gene; ds.
XX
OS   Homo sapiens.
XX
FH   Key              Location/Qualifiers
FH
FT   CDS              3000..56067
FT                    /tag= a
FT                    /product= "Serine/threonine kinase"
FT                    /note= "Contains 6 introns"
FT   variation        3000..3044
FT                    /tag= b
FT                    /standard_name= "Single nucleotide polymorphism"
FT                    /replace(941,T)
FT   variation        3000..3044
FT                    /tag= c
FT                    /standard_name= "Single nucleotide polymorphism"
FT                    /replace(2612,A)
FT   exon             3000..3044
FT                    /tag= d
FT                    /number= 1
FT   intron           3045..45393
FT                    /tag= e
FT                    /number= 1
FT   variation
FT                    /tag= f
FT                    /replace(5080,A)
FT                    /standard_name= "Single nucleotide polymorphism"
FT   variation
FT                    /tag= g
FT                    /replace(6598..9559,AAC,ACC)
FT                    /standard_name= "Single nucleotide polymorphism"
FT   variation
FT                    /tag= G
FT                    /replace(6983,G)
FT                    /standard_name= "Single nucleotide polymorphism"
FT   variation
FT                    /tag= h
FT                    /replace(9884..9886,AA)
FT                    /standard_name= "Single nucleotide polymorphism"
FT   variation
FT                    /tag= i
FT                    /replace(9884..9886,AA)
FT                    /standard_name= "Single nucleotide polymorphism"
```

```
Qy    599  AATGAA  604
            || ||
Db    316  ATTGGA  311

Qy    370  ------AAGAGTGAAACCCCATTCTCAAAAAAAAAAAAAAAAAAAAGCCAAAATAAA  317
                   |||||||||||||||||||||||| ||| |||||||||||||||||||| |||
Db    539  AGAGACAGAAATGAAAACTCTGTCTCAAAAACAAACAACAAAAACAACAACTATACATAAA  598
```

```
FT   variation
FT           /replace(12538,T)
FT           /*tag= j
FT   variation
FT           /replace(17707,C)
FT           /*tag= k
FT   variation
FT           /standard_name= "Single nucleotide polymorphism"
FT           /replace(18219..18220,AAT)
FT           /*tag= l
FT   variation
FT           /replace(19670,T)
FT           /*tag= m
FT   variation
FT           /standard_name= "Single nucleotide polymorphism"
FT           /replace(21153,T)
FT           /*tag= n
FT   variation
FT           /standard_name= "Single nucleotide polymorphism"
FT           /replace(24565..24567,CT)
FT           /*tag= o
FT   variation
FT           /standard_name= "Single nucleotide polymorphism"
FT           /replace(26604,A)
FT           /*tag= p
FT   variation
FT           /standard_name= "Single nucleotide polymorphism"
FT           /replace(27255,G)
FT           /*tag= q
FT   variation
FT           /standard_name= "Single nucleotide polymorphism"
FT           /replace(27399,C)
FT           /*tag= r
FT   variation
FT           /standard_name= "Single nucleotide polymorphism"
FT           /replace(28088,A)
FT           /*tag= s
FT   variation
FT           /standard_name= "Single nucleotide polymorphism"
FT           /replace(28734,A)
FT           /*tag= t
FT   variation
FT           /standard_name= "Single nucleotide polymorphism"
FT           /replace(29246..29247,TTA)
FT           /*tag= u
FT   variation
FT           /standard_name= "Single nucleotide polymorphism"
FT           /replace(29490,A)
FT           /*tag= v
FT   variation
FT           /standard_name= "Single nucleotide polymorphism"
FT           /replace(29934,C)
FT           /*tag= w
FT   variation
FT           /standard_name= "Single nucleotide polymorphism"
FT           /replace(34480,G)
FT           /*tag= x
FT   variation
FT           /standard_name= "Single nucleotide polymorphism"
FT           /replace(38812,C)
FT           /*tag= y
FT   variation
FT           /standard_name= "Single nucleotide polymorphism"
FT           /replace(40731,G)
FT           /*tag= z
FT   variation
FT           /standard_name= "Single nucleotide polymorphism"
FT           /replace(41303,A)
FT           /*tag= aa
FT   variation
FT           /standard_name= "Single nucleotide polymorphism"
FT           /replace(41305..41306, AAT)
FT           /*tag= ab
FT   variation
FT           /standard_name= "Single nucleotide polymorphism"
FT           /replace(41457,C)
FT           /*tag= ac
FT   variation
FT           /standard_name= "Single nucleotide polymorphism"
FT           /replace(43167,43169,AC,ATC)
FT           /*tag= ad
FT   variation
FT           /standard_name= "Single nucleotide polymorphism"
FT           /replace(43357,G)
FT           /*tag= ae
FT   exon
FT           /standard_name= "Single nucleotide polymorphism"
FT           /replace(45394..45525)
FT           /*tag= af
FT           /number= 2
FT   intron
FT           /45526..45761
FT           /*tag= ag
FT           /number= 2
FT   variation
FT           /replace(45664,C)

FT           /*tag= ah
FT   variation
FT           /standard_name= "Single nucleotide polymorphism"
FT           /replace(45762..45818)
FT           /*tag= ai
FT   exon
FT           /number= 3
FT           /45819..50154
FT           /*tag= aj
FT   intron
FT           /number= 4
FT           /replace(47549,C)
FT           /*tag= ak
FT   variation
FT           /standard_name= "Single nucleotide polymorphism"
FT           /replace(47908,A)
FT           /*tag= al
FT   exon
FT           /standard_name= "Single nucleotide polymorphism"
FT           /50155..50329
FT           /*tag= am
FT           /number= 3
FT   intron
FT           /50330..51076
FT           /*tag= an
FT           /number= 4
FT   exon
FT           /51077..51132
FT           /*tag= ao
FT           /number= 5
FT   intron
FT           /51133..52775
FT           /*tag= ap
FT           /number= 5
FT   variation
FT           /standard_name= "Single nucleotide polymorphism"
FT           /replace(52267,A)
FT           /*tag= aq
FT   exon
FT           /52776..52933
FT           /*tag= ar
FT           /number= 6
FT   intron
FT           /52934..55922
FT           /*tag= as
FT           /number= 6
FT   variation
FT           /replace(54654,C)
FT           /*tag= at
FT   variation
FT           /standard_name= "Single nucleotide polymorphism"
FT           /replace(54679,G)
FT           /*tag= au
FT   variation
FT           /standard_name= "Single nucleotide polymorphism"
FT           /replace(54693,C)
FT           /*tag= av
FT   variation
FT           /standard_name= "Single nucleotide polymorphism"
FT           /replace(54706,C)
FT           /*tag= aw
FT   variation
FT           /standard_name= "Single nucleotide polymorphism"
FT           /replace(54712,C)
FT           /*tag= ax
FT   variation
FT           /standard_name= "Single nucleotide polymorphism"
FT           /replace(54799,C)
FT           /*tag= ay
FT   variation
FT           /standard_name= "Single nucleotide polymorphism"
FT           /replace(54819,A)
FT           /*tag= az
FT   variation
FT           /standard_name= "Single nucleotide polymorphism"
FT           /replace(55499,T)
FT           /*tag= ba
FT   variation
FT           /standard_name= "Single nucleotide polymorphism"
FT           /55923..56067
FT           /*tag= bb
FT   exon
FT           /number= 7
FT           /replace(56825,A)
FT           /*tag= bc
FT   variation
FT           /replace(58871,A)
FT           /*tag= bd
FT   variation
FT           /standard_name= "Single nucleotide polymorphism"
XX
PN   US6340583-B1.
XX
PD   22-JAN-2002.
XX
```

Left column

```
PF   22-MAR-2001; 2001US-0813817.
XX
PR   22-MAR-2001; 2001US-0813817.
XX
PA   (PEKE ) PE CORP NY.
XX
PI   Yan C, Ketchum KA, Di Francesco V, Beasley EM;
XX
     WPI; 2002-224925/28.
DR   P-PSDB; ABB09972.
XX
```

```
Qy        1 CCTGCCACCATGCCTGGCTAATTTTCTTATTTTTATTTTTTAGTAGAGACGAGGTTTTGCCATGTTG   60
            ||||||||||| |||| || |||||||||||||||||||||||||||||||||||||||||||
Db    11786 CCGGCCCCCATACCCAGCTAATTTTTTATATTTTTTAGTAGAGATGGGGTTTTGCCATGTTG  11845

Qy       61 ACCAGGCTGGTCTCGAACTCTTGACCTCAGGTGATCCGCCTGCCTCAGCCTCCCAAAGTG  120
            ||| ||||| |||||||||| ||||||||| ||||||  || || |||||||||||
Db    11846 GCCAGGCTAGTCTCGAACTCCTGACCTCAAGTGATCCACCCGCCTCCCAAAGTG  11905

Qy      121 CTGGGATTATAGGCGTGCAGCCGCCGCGACCCAGC------CAACATTTTTAAATACTGAA  174
            ||||||||| ||| ||| |||| |||| ||   ||      ||| ||||||||||  |||
Db    11906 CTGGGATTACAGGTGTGAGCCACTGCACCTGGCCTGGAAGGAGTGATGTTAAATTTTTAAATATTTAAGGCA  11965

Qy      175 AAGTAGAGGAGGAATAGTTATAGTGTACCCCATTCACCCAGTTCAACAGCTGGT  234
            ||  ||  || || |    |  || |  ||| ||||||||||||||||||||||
Db    11966 AAACAAAAAAAACTTGACTGTGTCACTCTGTGTCTCTCCTACCTTGTTATACTTCCA  12025

Qy      235 GACATATTTATTTCTTCTATACCAGTACCGTACGTACTCTCCCCACTGGGGATTATTTAAGGCA  294
            ||| |  ||  |||    ||  |    |  |||| |||||||  ||||||||  |||| |||
Db    12026 CAACTTCCCAGTGTTCTTGGATAAAGCAACACCAAAAATCCTTAACTTGGCCAGGCGCGCGGTTGGCT  12085

Qy      295 AAACCCAGATGACATTTTATCCCTAAATACTTTAGATAAAGGTGTTCTTTGAAAAAAATC  354
            ||   |  ||| || ||  | ||||| |    ||||  || ||| ||||| |||||
Db    12086 CACACCTTATCATCT------CAGCACTTTGGGAGGCCGAGGCGGGCGGATCATGAAGTC  12138

Qy      355 ATAACCTTCAGACCACCAGCCTGCCCAACATGGTGAAACCCTGTCTGTTGTTACTAAAAAATACAAA  414
            || ||| || || ||||||| | || |||  ||| |||||||  | ||||| |||||  |||||||
Db    12139 AAGAGATTGAGACCATCCTGCCCAACATGGTGAAACCCCATCTCTACTAAAAATACAAAA  12198

Qy      415 ATTAGCTTGGCCATGGTGCGTGGGGCACCTGCCTGTAATCCCAGCTACTCAGGAAGCTGAGGCAGGA  474
            |||||||| | ||||| | | || |||||| | || |||||||| ||||  || || |||
Db    12199 ATTAGTTGGTCGTGGTGGTGCGTGCCTGTAGTCCCAGCTACTTGGGAGGCTGAGGCAGGA  12258

Qy      475 GAATCACTTGAATCCGGGAAGCAAGACGAATTGCAGTGAGCTGAGATTGCAGTGCGAGCCTGGG  534
            |||||||||||| |||||| |||| | || || ||||| |||| | |||||| || |||||||
Db    12259 GAATCACTTGAACCTGGGAGGCCAGAGGTTGCAGTGAGCCAAGATCACGCCACTGCACTCC  12318

Qy      535 CGACAGACAGAAAATGAAAACTCTGTCTCAAAAACAAACAAACAAAACAAAAA  584
            ||| | |  || |   || | ||||| |||||||  |||||||||||||||||
Db    12319 AGCCTGGTGACAGAGTAAGACTCCATCTCAAAAAAAAACAAACAAAACAAAAAAA  12368
```

```
RESULT 4
ABK83459/c
ID   ABK83459 standard; cDNA; 128978 BP.
XX
AC   ABK83459;
XX
DT   14-AUG-2002  (first entry)
XX
DE   Human cDNA differentially expressed in granulocytic cells #30.
XX
KW   Human; ss; granulocytic cell; DNA chip; bacterial infection;
KW   viral infection; parasitic infection; protozoal infection;
KW   fungal infection; sterile inflammatory disease; psoriasis;
KW   rheumatoid arthritis; glomerulonephritis; asthma; thrombosis;
KW   cardiac reperfusion injury; renal reperfusion injury; ARDS;
KW   adult respiratory distress syndrome; inflammatory bowel disease;
```

Right column

```
KW   Crohn's disease; ulcerative colitis; periodontal disease;
KW   granulocyte activation; chronic inflammation; allergy.
XX
OS   Homo sapiens.
XX
PN   WO20028999-A2.
DR   WPI; 2002-435328/46.
XX
PI   Beazer-Barclay Y, Weissman SM, Yamaga S, Vockley J;
XX
PA   (GENE-) GENE LOGIC INC.
XX
PR   03-OCT-2000; 2000US-2371189P.
XX
PD   03-OCT-2001; 2001WO-US30821.
XX
PN   11-APR-2002.
XX
PF   Detecting granulocyte activation by detecting differential expression
     of genes associated with granulocyte activation, which serves as
     diagnostic markers that is useful for monitoring disease states and
     drug toxicity -
XX
PS   Claim 1; SEQ ID No 30; 114pp; English.
XX
CC   The invention relates to detecting (M1) granulocyte (GC) activation
CC   (GCA), by detecting the level of expression of gene(s) (Gs) identified by
CC   DNA chip analysis as given in the specification, and comparing
CC   the expression level to an expression level in an unactivated
CC   GC, where differential expression of Gs is indicative of GCA.
CC   Also included are modulating (M2) GA by contacting GC with an agent
CC   that alters the expression of at least one gene in Gs; (2) screening (M3)
CC   for an agent capable of modulating GCA or an inflammation (especially
CC   chronic) in a tissue, an allergic response in a subject, exposure of a
CC   subject to a pathogen or sterile inflammatory disease using the
CC   gene expression profile; (3) detecting (M4) an inflammation (especially
CC   chronic) in a tissue, an allergic response in a subject, exposure of a
CC   subject to a pathogen or sterile inflammatory disease, by detecting the
CC   level of expression in a sample of the tissue of gene(s) from Gs, where
CC   the level of expression of the gene is indicative of inflammation,
CC   (4) treating (M5) an inflammation (especially chronic) or in a tissue,
CC   an allergic response in a subject, exposure of a subject to a pathogen
CC   or sterile inflammatory disease, by contacting a tissue having
CC   inflammation with an agent that modulates the expression of gene(s)
CC   from Gs in the tissue. M1 is useful for detecting GCA; M2 is useful for
CC   modulating GA; M3 is useful for screening an agent capable of modulating
CC   GCA preferably in an inflammation in a tissue; M4 is useful for
CC   detecting an inflammation (especially chronic) in a tissue, an allergic
CC   response in a subject, exposure of a subject to a pathogen or sterile
CC   inflammatory disease (e.g. psoriasis; rheumatoid arthritis,
CC   glomerulonephritis, asthma, thrombosis, cardiac reperfusion injury, renal
CC   reperfusion injury, ARDS, adult respiratory distress syndrome,
CC   inflammatory bowel disease, Crohn's disease, ulcerative colitis,
CC   periodontal disease; also bacterial infection, viral infection,
CC   parasitic infection, protozoal infection, fungal infection and M5 is
CC   useful for treating one of the above conditions. The present
CC   sequence represents a gene differentially expressed in granulocytes.
CC   Note: The sequence data for this patent did not form part
CC   of the printed specification, but was obtained from WIPO in
CC   format directly from WIPO at
CC   ftp.wipo.int/pub/published_pct_sequences.
XX
SQ   Sequence 128978 BP; 35134 A; 31020 C; 29358 G; 33466 T; 0 other;
XX
```

```
Query Match
Best Local Similarity   63.2%;  Pred. No. 3.6e-44;  DB 24;  Length 128978;
Matches 373;  Conservative  0;  Mismatches  204;  Indels  13;  Gaps  2;

Qy        1 CCTGCCACCATGCCTGGCTAATTTTCTTATTTTTATTTTTTAGTAGAGACGAGGTTTTGCCATGTTG   60
            ||||||||||| |||| || |||||||||||||||||||||||||||||||||||||||||||
Db    72729 CCGGCCCCCATACCCAGCTAATTTTTTATATTTTTTAGTAGAGATGGGGTTTTGCCATGTTG  72670
```

```
Qy         61 ACCAGGCTGTCTGGAACTCTTGACCTCAGGTGATCCGCCTTGCCTTCCCAAAGTG 120
              ||| ||||||||||||||||||| ||||||||||||||||||||||||||||||||
Db      72669 GCCAAGCTAGTCTCTGAACTCCTGACCTCAAGTGATCACCTGCCTTCCCAAAGTG 72610

Qy        121 CTGGGATTATAGGCGGTGAGCCGCCGCGCCCAGC-----CAACATTTTTAAATACTGAA 174
              |||||||||||||||||||||||| ||||||||||     || ||||||||||||||||
Db      72609 CTGGGATTACAGGTGTGAGCCACTGCCACCTGGCTGGAAGGAGTGGATTCTTAAAACAAA 72550

Qy        175 AAGTAGAGGGAATAGTTATATAGTGTGTACCCCATTTACCCATCACTCAGTTTCAACAGCTGGT 234
              |||||||||||||||||||||||||||||
Db      72549 AAACAAAAAAAAACTTGACTGTGTCACTCGTGTTCTCTGCTTGTTAAAAAAAAAA 72490

Qy        235 GACAATATTTTATTTCTTCTATACACAGTACCGTACTCTCCCCACTGGGATTATTTGTAAAGGCA 294
              |||| ||| ||| || ||||||| |||||||||||| ||||||||||||||||||
Db      72489 CAACTTCCCAGTGTTCTTGGATAAAGACCAAAAATCCTTAACTTGGCCAGGCGCGGT 72430

Qy        295 AAACCCAGATGACATTTTATCCCTAAAATACTTTAGATAAAGGTGTTCTTTGAAAAAAATC 354
              ||| ||
Db      72429 CACACCTATCATCT-----CACCACTTTGGGAGGCCGAGGCGCAGGCGAATCATGAAGTC 72377

Qy        355 ATAACCTCAGGACCAGCCTGGCCAACATGGTGAAACCCTGTCTGTACTAAAAATACAAAA 414
              ||| ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      72376 AAGAGATTGAGACCATCCTGGCCAACATGGTGAAACCCCATCCTACTAAAAATACAAAA 72317

Qy        415 ATTAGCTTGGCATGGTCGTCGGGCAACCTGTAATCCCAGCTACTCAGGAAGCTGAGGCAGGA 474
              |||||||||||||||||||||||| ||||||||||||||||||||||||||||||||||
Db      72316 ATTAGCTGGTCGTGGTGGCGTGGGCCTGTGTAGTCCCAGCTACTTGGGAGGCTGAGGCAGGA 72257

Qy        475 GAATCACTTGAATCCGGGAAGCACAGAATTGCAGTGAGCTGAGATTGCAGTCGAGGCTGGG 534
              |||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      72256 GAATCACTTGAACCTGGGAGGCAGAGGTTGCAGTGAGCCCGAGATCACGCCCACTGCACTCC 72197

Qy        535 CGACAGAGACAGAAATGAAAACTCTGTCTCAAAAACAAAACAAAACAAAAAAA 584
              ||| ||||||||||||||||||||| ||| |||||| |||  ||||||||||
Db      72196 AGCCTGGTGACAGAGTAAGACTCCATCTCAAAAAAAAAAAAAAAAAAAAAAAA 72147
```

RESULT 5
AAL05122/c
ID   AAL05122 standard; DNA; 13216 BP.
XX
AC   AAL05122;
XX
DT   21-NOV-2001 (first entry)
XX
DE   Human reproductive system related antigen DNA SEQ ID NO: 7810.
XX
KW   Human; reproductive system related antigen; reproductive system disorder;
KW   cancer; gene therapy; ds.
XX
OS   Homo sapiens.
XX
PN   WO200155320-A2.
XX
PD   02-AUG-2001.
XX
PF   17-JAN-2001; 2001WO-US01339.
XX
PR   31-JAN-2000; 2000US-0179065.
PR   04-FEB-2000; 2000US-0180628.
PR   24-FEB-2000; 2000US-0184664.
PR   02-MAR-2000; 2000US-0186350.
PR   16-MAR-2000; 2000US-0189874.
PR   17-MAR-2000; 2000US-0190076.
PR   18-APR-2000; 2000US-0194123.
PR   19-MAY-2000; 2000US-0205515.
PR   07-JUN-2000; 2000US-0209467.
PR   28-JUN-2000; 2000US-0214886.
PR   30-JUN-2000; 2000US-0215135.
PR   07-JUL-2000; 2000US-0216647.
PR   07-JUL-2000; 2000US-0216880.
PR   11-JUL-2000; 2000US-0217487.
PR   11-JUL-2000; 2000US-0217496.
PR   14-JUL-2000; 2000US-0218290.
PR   26-JUL-2000; 2000US-0220963.
PR   26-JUL-2000; 2000US-0220964.
PR   14-AUG-2000; 2000US-0224518.
PR   14-AUG-2000; 2000US-0224519.
PR   14-AUG-2000; 2000US-0225213.
PR   14-AUG-2000; 2000US-0225214.
PR   14-AUG-2000; 2000US-0225266.
PR   14-AUG-2000; 2000US-0225267.
PR   14-AUG-2000; 2000US-0225758.
PR   14-AUG-2000; 2000US-0225759.
PR   18-AUG-2000; 2000US-0226279.
PR   22-AUG-2000; 2000US-0266868.
PR   22-AUG-2000; 2000US-0266681.
PR   22-AUG-2000; 2000US-0227182.
PR   23-AUG-2000; 2000US-0227009.
PR   30-AUG-2000; 2000US-0228924.
PR   01-SEP-2000; 2000US-0229287.
PR   01-SEP-2000; 2000US-0229343.
PR   01-SEP-2000; 2000US-0229344.
PR   05-SEP-2000; 2000US-0229345.
PR   05-SEP-2000; 2000US-0229509.
PR   05-SEP-2000; 2000US-0229513.
PR   06-SEP-2000; 2000US-0230437.
PR   06-SEP-2000; 2000US-0230438.
PR   08-SEP-2000; 2000US-0231242.
PR   08-SEP-2000; 2000US-0231243.
PR   08-SEP-2000; 2000US-0231244.
PR   08-SEP-2000; 2000US-0231413.
PR   08-SEP-2000; 2000US-0231414.
PR   08-SEP-2000; 2000US-0232080.
PR   08-SEP-2000; 2000US-0232081.
PR   12-SEP-2000; 2000US-0231968.
PR   14-SEP-2000; 2000US-0232397.
PR   14-SEP-2000; 2000US-0232398.
PR   14-SEP-2000; 2000US-0232399.
PR   14-SEP-2000; 2000US-0232400.
PR   14-SEP-2000; 2000US-0232401.
PR   14-SEP-2000; 2000US-0233063.
PR   14-SEP-2000; 2000US-0233064.
PR   14-SEP-2000; 2000US-0233065.
PR   21-SEP-2000; 2000US-0234223.
PR   21-SEP-2000; 2000US-0234274.
PR   25-SEP-2000; 2000US-0234997.
PR   25-SEP-2000; 2000US-0234998.
PR   26-SEP-2000; 2000US-0235484.
PR   27-SEP-2000; 2000US-0235834.
PR   27-SEP-2000; 2000US-0235836.
PR   29-SEP-2000; 2000US-0236327.
PR   29-SEP-2000; 2000US-0236367.
PR   29-SEP-2000; 2000US-0236368.
PR   29-SEP-2000; 2000US-0236369.
PR   29-SEP-2000; 2000US-0236370.
PR   02-OCT-2000; 2000US-0236802.
PR   02-OCT-2000; 2000US-0237037.
PR   02-OCT-2000; 2000US-0237038.
PR   02-OCT-2000; 2000US-0237039.
PR   02-OCT-2000; 2000US-0237040.
PR   13-OCT-2000; 2000US-0239935.
PR   13-OCT-2000; 2000US-0239937.
PR   20-OCT-2000; 2000US-0240960.
PR   20-OCT-2000; 2000US-0241221.
PR   20-OCT-2000; 2000US-0241785.
PR   20-OCT-2000; 2000US-0241786.
PR   20-OCT-2000; 2000US-0241787.
PR   20-OCT-2000; 2000US-0241808.
PR   20-OCT-2000; 2000US-0241809.
PR   20-OCT-2000; 2000US-0241826.
PR   01-NOV-2000; 2000US-0244617.

```
PR   08-NOV-2000;  2000US-0246474.
PR   08-NOV-2000;  2000US-0246475.
PR   08-NOV-2000;  2000US-0246476.
PR   08-NOV-2000;  2000US-0246477.
PR   08-NOV-2000;  2000US-0246478.
PR   08-NOV-2000;  2000US-0246523.
PR   08-NOV-2000;  2000US-0246524.
PR   08-NOV-2000;  2000US-0246525.
PR   08-NOV-2000;  2000US-0246526.
PR   08-NOV-2000;  2000US-0246527.
PR   08-NOV-2000;  2000US-0246528.
PR   08-NOV-2000;  2000US-0246532.
PR   08-NOV-2000;  2000US-0246609.
PR   08-NOV-2000;  2000US-0246610.
PR   08-NOV-2000;  2000US-0246611.
PR   17-NOV-2000;  2000US-0246613.
PR   17-NOV-2000;  2000US-0249207.
PR   17-NOV-2000;  2000US-0249208.
PR   17-NOV-2000;  2000US-0249209.
PR   17-NOV-2000;  2000US-0249210.
PR   17-NOV-2000;  2000US-0249211.
PR   17-NOV-2000;  2000US-0249212.
PR   17-NOV-2000;  2000US-0249213.
PR   17-NOV-2000;  2000US-0249214.
PR   17-NOV-2000;  2000US-0249215.
PR   17-NOV-2000;  2000US-0249216.
PR   17-NOV-2000;  2000US-0249217.
PR   17-NOV-2000;  2000US-0249218.
PR   06-DEC-2000;  2000US-0251479.
PR   05-DEC-2000;  2000US-0251856.
PR   05-DEC-2000;  2000US-0251868.
PR   08-DEC-2000;  2000US-0251869.
PR   08-DEC-2000;  2000US-0251989.
PR   08-DEC-2000;  2000US-0251990.
PR   11-DEC-2000;  2000US-0254097.
PR   05-JAN-2001;  2001US-0259678.

XX
PA   (HUMA-) HUMAN GENOME SCI INC.
XX
PI   Rosen CA,  Barash SC,  Ruben SM;
XX
DR   WPI: 2001-465570/50.
XX
PT   Isolated nucleic acid molecule encoding a reproductive system antigen
     is used in preventing, treating or ameliorating a medical condition -
XX
PS   Disclosure; SEQ ID NO 7810; 1297pp + Sequence Listing; English.
XX
CC   The present invention provides the protein and coding sequences of a
CC   number of human reproductive system related antigens. These can be used
CC   in the prevention and treatment of reproductive system disorders,
CC   including cancer. The present sequence is a genomic sequence encoding a
CC   protein of the invention.
XX
SQ   Sequence 13216 BP; 3041 A; 3400 C; 3243 G; 3532 T; 0 other;
```

Query Match   21.1%;   Score 211; DB 22;   Length 13216;
Best Local Similarity   62.1%;   Pred. No. 8e-43;
Matches 361; Conservative   0; Mismatches 200; Indels 20; Gaps   1;

```
Qy          4 GCCACCACATGCCTGGCTAATTTCTTATTTTTTAGTAGAGACGAGGTTTGCCATGTTGACC 63
              |||||||  ||   ||||||       ||||||||||||||||||||||||||||||||||
Db       4181 GCCACCACACGTGGCTAATTTGTATTTTTAGTAGAGACAGGGTGCTCACCTTGTTGGCC 4122

Qy         64 AGGCTGGTCTCGAACTCTTGACCTCAGGTGATCCGCGCCTCAGCCTCCCAAAGTGCTG 123
              |||||||||||||||| ||| ||||||  |||||||  |||||||||||||||||||||||
Db       4121 AGGCTGGTCTTGAACTCCGGACCTCAGTGATCCACCCGCCTCGGCCTCCCAAAGTGCTG 4062

Qy        124 GGATTATACGGCGTGAGCCGCCGCACCCACCCACACATTTTTTAAATACTGAAAAGTAGAG 183
              |||||||| |||||||  ||||||  ||| || || || |||   |||||   ||| ||||
Db       4061 GAATTACAGGCGTGAGCCACCGCACCCGCCCTGTAATCCCAGCTACTTTGGGAGGCCGAGGC 4002

Qy        184 GAATAGTTATTAGTGTACCCCATTTACCCATCACTCAGTTTCAACAGCTGGTGTGACATATTT 243
              || ||||  |||  || || | | ||||| ||  || | ||| ||| |  | || || | |
Db       4001 AGATGGTTCACTTGCGCCCAGGAGTTCGAGACCAGCCTGGGCAACATGGAGAAACTCTCGT 3942

Qy        244 ATTTCTTCTTATACCAGTACCGTACTCTCCCCACTGGGGATTATTTTGGGCAAAACCCAGA 303
              | ||   | ||||| |||| || |  | || |||| ||| |||| ||  |||||  |||
Db       3941 CTGTACAAAAATAAATTAGCCGGGTGGGACAGTCG---------------------CTCA 3902

Qy        304 TGACATTTTATCCCTAAAATACTTTAGATAAAGGTGTTCTTTGAAAAAAATCATAACCTCA 363
              |  ||  ||| || |||||  ||| | |  ||||| || || | |||||||||||||||||
Db       3901 TGCCTGTAATCCTAACACTTTGGGAGGTCGAGGCGGGCGGATCATTGAAGTCAGGAGTTCG 3842

Qy        364 GGACCAGCCTGGCCAACATGGTGAAACCCTGTCTCTACTAAAAATACAAAAATTAGCTG 423
              || ||||||| |||||||||  |||||| |||   || || ||  ||||||||||||  |
Db       3841 AGACCAGCCTGACCAACATGGTGAAACCCATCGTATTAAAAATATAAAAATTAGCCAG 3782

Qy        424 GCATGGTGGCGTGTGGGCGACCTGTAATCCCAGCTACTCAGGAGGACTGAGGCAGGAGAATCACTT 483
              |||||||||| |||  || |||| ||||||||||||||  |||||| ||||| |||||||||| ||
Db       3781 GCACGGTAGTGCACCCGTGTAATCCCAGCTACTCGGGAGGCTGAGGCAGGAGAATCTCTT 3722

Qy        484 GAATCCGGGAAGCAGAGAGTTGCAGTGAGCTGAGATTGCAGTCGAGCCTGGGCGACAGAGA 543
              |||| |||| ||| ||| |||||||||| ||||| |||| ||| ||| ||||| || |||
Db       3721 GAACCCGGGAGGTGGAGTTTGCAGTGAGCCGAGATCGCACCACTGCACTCCAGCCTGGGC 3662

Qy        544 CAGAAATGAAACTCTGTTCTCAAAAACAAACAAACAAACAAAAAAA 584
              ||| | |||||  | |  |||||||||||| |||||||||||||||
Db       3661 GACAAAGCAAGACTCCAGAAAAAAAAAAAAAAGAAAAAGAAAAAA 3621
```

```
RESULT 6
ABL98014/c
ID   ABL98014 standard; DNA; 13216 BP.
XX
AC   ABL98014;
XX
DT   21-JUN-2002 (first entry)
XX
PN   WO200155317-A2.
XX
PD   02-AUG-2001.
XX
OS   Homo sapiens.
XX
DE   Human testicular antigen encoding DNA fragment SEQ ID NO: 2666.
XX
KW   Human; testicular antigen; testes; cancer; metastasis; immune disorder;
KW   reproductive system disorder; urinary system disorder; gene therapy;
KW   cardiovascular disorder; respiratory disorder; neurological disorder;
KW   gastrointestinal disease; infection; cytostatic; gene; ds.
XX
PF   17-JAN-2001;  2001WO-US01329.
XX
PR   31-JAN-2000;  2000US-0179065.
PR   04-FEB-2000;  2000US-0180628.
PR   24-FEB-2000;  2000US-0184664.
PR   02-MAR-2000;  2000US-0186350.
PR   16-MAR-2000;  2000US-0189874.
PR   17-MAR-2000;  2000US-0190076.
PR   18-APR-2000;  2000US-0198123.
PR   19-MAY-2000;  2000US-0205515.
PR   07-JUN-2000;  2000US-0209467.
PR   28-JUN-2000;  2000US-0214886.
```

```
PR   30-JUN-2000; 2000US-0215135.          PR   20-OCT-2000; 2000US-0241787.
PR   07-JUL-2000; 2000US-0216647.          PR   20-OCT-2000; 2000US-0241808.
PR   07-JUL-2000; 2000US-0216880.          PR   20-OCT-2000; 2000US-0241809.
PR   11-JUL-2000; 2000US-0217487.          PR   20-OCT-2000; 2000US-0241926.
PR   11-JUL-2000; 2000US-0217496.          PR   01-NOV-2000; 2000US-0244617.
PR   14-JUL-2000; 2000US-0218290.          PR   08-NOV-2000; 2000US-0246474.
PR   26-JUL-2000; 2000US-0220963.          PR   08-NOV-2000; 2000US-0246475.
PR   26-JUL-2000; 2000US-0220964.          PR   08-NOV-2000; 2000US-0246476.
PR   14-AUG-2000; 2000US-0224518.          PR   08-NOV-2000; 2000US-0246477.
PR   14-AUG-2000; 2000US-0224519.          PR   08-NOV-2000; 2000US-0246478.
PR   14-AUG-2000; 2000US-0245213.          PR   08-NOV-2000; 2000US-0246523.
PR   14-AUG-2000; 2000US-0225214.          PR   08-NOV-2000; 2000US-0246524.
PR   14-AUG-2000; 2000US-0255266.          PR   08-NOV-2000; 2000US-0246525.
PR   14-AUG-2000; 2000US-0255267.          PR   08-NOV-2000; 2000US-0246526.
PR   14-AUG-2000; 2000US-0225268.          PR   08-NOV-2000; 2000US-0246527.
PR   14-AUG-2000; 2000US-0225270.          PR   08-NOV-2000; 2000US-0246528.
PR   14-AUG-2000; 2000US-0225447.          PR   08-NOV-2000; 2000US-0246532.
PR   14-AUG-2000; 2000US-0225757.          PR   08-NOV-2000; 2000US-0246609.
PR   14-AUG-2000; 2000US-0225758.          PR   08-NOV-2000; 2000US-0246610.
PR   14-AUG-2000; 2000US-0225759.          PR   08-NOV-2000; 2000US-0246611.
PR   18-AUG-2000; 2000US-0226279.          PR   08-NOV-2000; 2000US-0246613.
PR   22-AUG-2000; 2000US-0226681.          PR   17-NOV-2000; 2000US-0249207.
PR   22-AUG-2000; 2000US-0226868.          PR   17-NOV-2000; 2000US-0249208.
PR   22-AUG-2000; 2000US-0227182.          PR   17-NOV-2000; 2000US-0249209.
PR   23-AUG-2000; 2000US-0227009.          PR   17-NOV-2000; 2000US-0249210.
PR   30-AUG-2000; 2000US-0228924.          PR   17-NOV-2000; 2000US-0249211.
PR   01-SEP-2000; 2000US-0229287.          PR   17-NOV-2000; 2000US-0249212.
PR   01-SEP-2000; 2000US-0229343.          PR   17-NOV-2000; 2000US-0249213.
PR   01-SEP-2000; 2000US-0229344.          PR   17-NOV-2000; 2000US-0249214.
PR   01-SEP-2000; 2000US-0229345.          PR   17-NOV-2000; 2000US-0249215.
PR   05-SEP-2000; 2000US-0229509.          PR   17-NOV-2000; 2000US-0249216.
PR   05-SEP-2000; 2000US-0229513.          PR   17-NOV-2000; 2000US-0249217.
PR   06-SEP-2000; 2000US-0230437.          PR   17-NOV-2000; 2000US-0249218.
PR   06-SEP-2000; 2000US-0230438.          PR   17-NOV-2000; 2000US-0249244.
PR   08-SEP-2000; 2000US-0231242.          PR   17-NOV-2000; 2000US-0249245.
PR   08-SEP-2000; 2000US-0231243.          PR   17-NOV-2000; 2000US-0249264.
PR   08-SEP-2000; 2000US-0231244.          PR   17-NOV-2000; 2000US-0249265.
PR   08-SEP-2000; 2000US-0231413.          PR   17-NOV-2000; 2000US-0249297.
PR   08-SEP-2000; 2000US-0231414.          PR   17-NOV-2000; 2000US-0249299.
PR   08-SEP-2000; 2000US-0231968.          PR   17-NOV-2000; 2000US-0249300.
PR   12-SEP-2000; 2000US-0232080.          PR   01-DEC-2000; 2000US-0250160.
PR   08-SEP-2000; 2000US-0232081.          PR   01-DEC-2000; 2000US-0250391.
PR   14-SEP-2000; 2000US-0232397.          PR   05-DEC-2000; 2000US-0251030.
PR   14-SEP-2000; 2000US-0232398.          PR   05-DEC-2000; 2000US-0251988.
PR   14-SEP-2000; 2000US-0232399.          PR   05-DEC-2000; 2000US-0256719.
PR   14-SEP-2000; 2000US-0232400.          PR   06-DEC-2000; 2000US-0251479.
PR   14-SEP-2000; 2000US-0232401.          PR   08-DEC-2000; 2000US-0251856.
PR   14-SEP-2000; 2000US-0233063.          PR   08-DEC-2000; 2000US-0251868.
PR   14-SEP-2000; 2000US-0233064.          PR   08-DEC-2000; 2000US-0251869.
PR   14-SEP-2000; 2000US-0233065.          PR   08-DEC-2000; 2000US-0251889.
PR   21-SEP-2000; 2000US-0234223.          PR   08-DEC-2000; 2000US-0251990.
PR   21-SEP-2000; 2000US-0234274.          PR   11-DEC-2000; 2000US-0254097.
PR   25-SEP-2000; 2000US-0234997.          PR   05-JAN-2001; 2001US-0259678.
PR   25-SEP-2000; 2000US-0234998.
PR   26-SEP-2000; 2000US-0235484.
PR   27-SEP-2000; 2000US-0235834.
PR   29-SEP-2000; 2000US-0235836.
PR   29-SEP-2000; 2000US-0236327.
PR   29-SEP-2000; 2000US-0236367.
PR   29-SEP-2000; 2000US-0236368.
PR   29-SEP-2000; 2000US-0236369.
PR   29-SEP-2000; 2000US-0236370.
PR   02-OCT-2000; 2000US-0236802.
PR   02-OCT-2000; 2000US-0237037.
PR   02-OCT-2000; 2000US-0237038.
PR   02-OCT-2000; 2000US-0237039.
PR   02-OCT-2000; 2000US-0237040.
PR   13-OCT-2000; 2000US-0239935.
PR   13-OCT-2000; 2000US-0239937.
PR   20-OCT-2000; 2000US-0240960.
PR   20-OCT-2000; 2000US-0241221.
PR   20-OCT-2000; 2000US-0241785.
PR   20-OCT-2000; 2000US-0241786.
```

XX

PA   (HUMA-) HUMAN GENOME SCI INC.

XX

PI   Rosen CA,   Barash SC,   Ruben SM;

XX

DR   WPI; 2001-483232/52.

XX

PT   Nucleic acids encoding 973 human testicular antigen polypeptides,
PT   useful for preventing, diagnosing and/or treating testicular cancer -

XX

PS   Disclosure; SEQ ID NO 2666; 766pp; English.

XX

CC   The present invention provides the protein and coding sequences of 973
CC   human testicular antigens, and fragments of their genomic sequences. The
CC   sequences can be used in the treatment of cardiovascular, urinary system,
CC   reproductive system, immune, respiratory, neurological and
CC   gastrointestinal disorders, infections, and particularly cancer,
CC   especially testicular cancers. The present sequence is a DNA encoding a
CC   protein fragment of the invention.

XX

```
SQ   Sequence 13216 BP; 3041 A; 3400 C; 3243 G; 3532 T; 0 other;

Query Match              21.1%;  Score 211;  DB 23;  Length 13216;
Best Local Similarity    62.1%;  Pred. No. 8e-43;
Matches 361;  Conservative  0;  Mismatches 200;  Indels 20;  Gaps  1;

Qy      4 GCCACCACGCCTGGCTAATTTTTTTTTATTTTTAGTAGAGACGAGGTTTTGCCATGTTGACC 63
          |||||| ||||||||||||| ||||||||||||||||||||||||||| ||| ||||| |||
Db   4181 GCCACCACGCCTGGCTAATTTTGTATTTTTAGTAGAGACAGGGTGTCACCTTGTTGGCC 4122

Qy     64 AGGCTGGTCTCGAACTCTTGACCTCAGGTGATCCGCCGCCTTCAGCCTCCCAAAGTGCTG 123
          ||| ||| ||||||||| |||||||||||||||||||||||||| ||||||||||||||||
Db   4121 AGGGTGGTCTTGAACTCCGAGACCTCAGGTGATCCACCTGCCTTGGGCCTCCCAAAGTGCTG 4062

Qy    124 GGATTATAGGCGTGAGCCGCCGCGCCCAGCCAACATTTTTAAATACTGAAAAGTAGAGG 183
          |||||||||||||||||||||||||| ||||||||||||||||||||||||||||||||
Db   4061 GGATTACAGGCGTGAGCCACCGCACCCTGCCTGTAATCCCAGCACTTTGGAGGCCCGAGGC 4002

Qy    184 GAATAGTTATAGTGTACCCCATTACCATCACTCAGTTTCAACAGCTGGTGACTATATTT 243
          ||| || ||||| ||||||||||  ||||||| |||||||||||||||||||||||||||
Db   4061 AGATGGTTCACTTGGGCCCAGGAGTTCGAGATCAGCCTGGCAACATGGAGAACTCTGT 3942

Qy    244 ATTTCTTCTATACCAGTACCTACCCTGGGATTATTTTTTAGGCAAAACCCAGA 303
          ||| | |||  ||||||| ||||||||||||||||||||||||||||||
Db   3941 CTGTACAAAAAAATAAATTAGGGGCCGGCTCGGCACAGTGG-----------------CTCA 3902

Qy    304 TGACATTTTATCCCTAAATACTTTACATAAAAGGTGTTCTTTGAAAAAAATCATAACCTCA 363
          |||||||| |||||||||||||   |||  || |||  |
Db   3901 TGCCTGTAATCCTAACACTTTGGGAGGCTGAGGCGGATCATGAGGTCAGGAGTTCG 3842

Qy    364 GGACCAGCCTGGCCAACATGGTGAAAACCCTGTCTCTGTACTAAAAATACAAAAAATTAGCTTG 423
          ||  || ||| |||||| ||||||| || |||||||||||||||||  |||||||||||
Db   3841 AGACCAGCCTGACCAACATGTGAAACCCCATCTGTTATTAAAATATAAAAATTAGCCAG 3782

Qy    424 GCATGGTGGTCGTGGGCACCTGTAATCCCAGCTACTCAGGAGGCTGAGGCAGGAGAATCACTT 483
          ||||| |||| ||  ||||||||||||||||||||||||||||||||||||| || |
Db   3781 GCACGGTAGTCCACGCCCTGTAATCCCAGCTACTCAGGAGGCTAAGGCAGGAGAATCTCTT 3722

Qy    484 GAATCCGGGAGGCAGAGGATTGCAGTGAGCTGAGATTGCAGTGAGCCGAGATTGTGCCGACAGAGA 543
          |||| ||||||||||||||||||||||||||||||||| || || || ||||||||
Db   3721 GAACCCGGGAGGCGGAGTTTGGCAGTGAGCCGAGATCGCCACCACTGCACTCCAGCCTGGGC 3662

Qy    544 CAGAAATGAAACTCTGTCTCAAAAACAAAACAAAACAAAAAAA 584
          ||| |||||||||||||||||||||||
Db   3661 GACAAAGCAAGACTCCAGAAAAGAAAAGAAAAGAAAAAAA 3621
```

Human osteoblast differentiation related cDNA SEQ ID NO 93.

```
RESULT 7
AB088186
ID   AB088186 standard; cDNA; 169739 BP.
XX
AC   AB088186;
XX
DT   18-SEP-2002 (first entry)
XX
DE   Human osteoblast differentiation related cDNA SEQ ID NO 93.
XX
KW   Human; osteoblast; stem cell differentiation; bone tissue deposition;
KW   osteoporosis; osteopathic; ss.
XX
OS   Homo sapiens.
XX
PN   WO200250301-A2.
XX
PD   27-JUN-2002.
XX
PF   18-DEC-2001; 2001WO-US48276.
PR   18-DEC-2000; 2000US-2558882P.
PR   24-APR-2001; 2001US-2856919P.
XX
PA   (GENE-) GENE LOGIC INC.
PA   (PROC ) PROCTER & GAMBLE CO.
```

PA   (PROC) PROCTER & GAMBLE CO.
XX
PI   Ji D, Axelrod DW, Cook JS, Jaiswal N, Einstein R, Houghton A;
PI   Mertz L;
XX
DR   WPI; 2002-557663/59.
XX
PT   Use of genes and their expression profiles associated with osteoblast
PT   differentiation for screening modulators bone formation, for diagnosing
PT   or treating e.g. osteoporosis, or as markers for the differentiation
PT   process.
XX
PS   Claim 1; SEQ ID NO 93; 78pp + Sequence Listing; English.
XX
CC   The invention relates to genes and their expression profiles are used
CC   for:
CC   (a) screening modulators of precursor stem cell differentiation into
CC   osteoblasts, or bone tissue deposition;
CC   (b) diagnosing abnormal deposition of bone tissue, abnormal rate of
CC   osteoblast formation or osteoporosis; or
CC   (c) treating or monitoring treatment of the conditions cited in (b), or
CC   monitoring the progression of bone tissue deposition.
CC   Specific conditions include postmenopausal osteoporosis, glucocorticoid
CC   drug-induced abnormalities in bone formation or bone loss, conditions
CC   that involve altered bone metabolism (e.g. osteopenia, osteodystrophy,
CC   osteoporosis), skeletal disease linked to breast cancer, mastocytosis,
CC   Fanconi syndrome or fibrous dysplasia. The present sequence is that of an
CC   osteoblast differentiation associated cDNA marker of the invention.
CC   Note: The sequence data for this patent did not form part of the printed
CC   specification, but was obtained in electronic format directly from WIPO
CC   at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ   Sequence 169739 BP; 49809 A; 35660 C; 35715 G; 48555 T; 0 other;

```
Query Match              21.0%;  Score 209.8;  DB 24;  Length 169739;
Best Local Similarity    62.3%;  Pred. No. 3.5e-42;
Matches 362;  Conservative  0;  Mismatches 217;  Indels  2;  Gaps  2;

Qy      4 GCCACCACGCCTGGCTAATTTTTTTTTATTTTTAGTAGAGACGAGGTTTTGCCATGTTGACC 63
          |||||| ||||||||||||| |||||||||||||||||||||||||||| ||| ||||| |||
Db 138055 GCCACCAGCCCAGATAATTTT-GTATTTTTAGTAGAGACG-GGTTTCTCCATGTTGGTC 137994

Qy     64 AGGCTGGTCTCGAACTCTTGACCTCAGGTGATCCGCCGCCTTCAGCCTCCCAAAGTGCTG 123
          ||||||||| |||||||||||||||||||||||||| ||||||||||||||||||||||||
Db 137995 AGGCTGGTGTCAAACTCCTGACCTCAGGGGATCCGCCTGCCTCAGCCCCTAAAGTGCTA 138054

Qy    124 GGATTATAGGCGTGAGCCGCCGCGCCCAGCCAACATTTTTAAATACTGAAAAGTAGAGG 183
          |||| || ||||||||||||||||||||||||| ||||||||||||||||||||||||||
Db 138055 GGATTACTGGCATGAGCCACCTGCACCTCACCCGCGAAAAATAAATATTTTATAACCTGAGGCAA 138114

Qy    184 GAATAGTTATAGTGTACCCCATTACCATCACTCAGTTTCAACAGCTGGTGACATATTT 243
          ||||||||| |||| |||||| ||||||||| ||| |||| ||||||||||||||||
Db 138115 GCTAGAAGACAGTATAACCAGAGGGATTTCTAGAGGCACTATCCATTACGGTAGCCACTA 138174

Qy    244 ATTTCTTCTATACCAGTACCTACCCTGTCTGTTGTACTAAAAATACAAAAAATTAGCTTG 303
          |||||  | |||| ||||||||||| ||||||||||||||||||||||
Db 138175 GCTACACGTGCCTATTTAATTATAAAATATAAAAATTCAGCGGGCGTGGTCTCA 138234

Qy    304 TGACATTTTATCCCTAAATACTTTAGAGTAAAGGTGTTCTTTGAAAAAAATCATAACCTCA 363
          |||||||||| | |||||||||||||| ||||||||||||  |||||||||
Db 138235 AGCCTGTAATCCCAGCACTTTGGAAGGCCAAGGTGGGGGATCACAAGGTCAAAAAGATCA 138294

Qy    364 GGACCAGCCTGGCCAACATGGTGAAAACCCTGTCTGTTACTAAAAATACAAAAAATTAGCTTG 423
          ||||||| ||| |||||| ||||||||||| ||||| |||||||||||||
Db 138295 AGACCATCCTGGCCAACATGGTGAAATGCCGTCTTTACTAAAAATACAAAAAATTAGCTGG 138354

Qy    424 GCATGGTGGTCGTGGGCACCTGTAATCCCAGCTACTCAGGAGGCTGAGGCAGGAGAATCACTT 483
          ||||||| ||| || || ||||| ||||||||||||||||||||||||||||||||||||
Db 138355 GCATGGTGGCGCGTGCCTGTAATCCCAGCTACTCGGGAGGCTGAGGCAGGAGAATGGCTT 138414

Qy    484 GAATCCGGGAGGCAGAGGATTGCAGTGAGCTGAGATTGCGCCACTCCAGCCTGGGTG 543
          |||| |||||| | ||||||||||||||| ||||| || || || ||||||||||
Db 138415 GAACCTGGGAGGCGGAGGTTGCACTGAGCTGAGATTGTGCCACTGCACTCCAGCCTGGGTG 138474
```

```
Qy     544 CAGAAATGAAACTCTGTCTCAAAAACAAACAAACAAAAAAA 584
           ||||||||||||| ||| ||| | | | |
Db 138475 ACAGAGCCAGACTCTGTCTCAAAAAAAAAAAAAAAAGATATA 138515
```

RESULT 8
```
ID   AAL05121/c standard; DNA; 29163 BP.
XX
AC   AAL05121;
XX
DT   21-NOV-2001 (first entry)
XX
DE   Human reproductive system related antigen DNA SEQ ID NO: 7809.
XX
KW   Human; reproductive system related antigen; reproductive system disorder;
KW   cancer; gene therapy; ds.
XX
OS   Homo sapiens.
XX
PN   WO200155320-A2.
XX
PD   02-AUG-2001.
XX
PF   17-JAN-2001; 2001WO-US01339.
XX
PR   31-JAN-2000;  2000US-0179065.
PR   04-FEB-2000;  2000US-0180628.
PR   24-FEB-2000;  2000US-0184664.
PR   02-MAR-2000;  2000US-0186350.
PR   16-MAR-2000;  2000US-0189874.
PR   17-MAR-2000;  2000US-0190076.
PR   18-APR-2000;  2000US-0198123.
PR   19-MAY-2000;  2000US-0205515.
PR   07-JUN-2000;  2000US-0209467.
PR   28-JUN-2000;  2000US-0214886.
PR   30-JUN-2000;  2000US-0215135.
PR   07-JUL-2000;  2000US-0216647.
PR   07-JUL-2000;  2000US-0216880.
PR   11-JUL-2000;  2000US-0217487.
PR   14-JUL-2000;  2000US-0217496.
PR   26-JUL-2000;  2000US-0218290.
PR   26-JUL-2000;  2000US-0220963.
PR   14-AUG-2000;  2000US-0220964.
PR   14-AUG-2000;  2000US-0224518.
PR   14-AUG-2000;  2000US-0224519.
PR   07-AUG-2000;  2000US-0225213.
PR   14-AUG-2000;  2000US-0225214.
PR   14-AUG-2000;  2000US-0225266.
PR   14-AUG-2000;  2000US-0225267.
PR   14-AUG-2000;  2000US-0225268.
PR   14-AUG-2000;  2000US-0225270.
PR   14-AUG-2000;  2000US-0225447.
PR   14-AUG-2000;  2000US-0225457.
PR   14-AUG-2000;  2000US-0225758.
PR   14-AUG-2000;  2000US-0225759.
PR   18-AUG-2000;  2000US-0226279.
PR   22-AUG-2000;  2000US-0226681.
PR   22-AUG-2000;  2000US-0226868.
PR   22-AUG-2000;  2000US-0227182.
PR   23-AUG-2000;  2000US-0227009.
PR   30-AUG-2000;  2000US-0228924.
PR   01-SEP-2000;  2000US-0229287.
PR   01-SEP-2000;  2000US-0229343.
PR   01-SEP-2000;  2000US-0229344.
PR   01-SEP-2000;  2000US-0229345.
PR   05-SEP-2000;  2000US-0229509.
PR   05-SEP-2000;  2000US-0229513.
PR   06-SEP-2000;  2000US-0230437.
PR   06-SEP-2000;  2000US-0230438.
PR   08-SEP-2000;  2000US-0231242.
PR   08-SEP-2000;  2000US-0231243.
PR   08-SEP-2000;  2000US-0231244.
PR   08-SEP-2000;  2000US-0231413.
PR   08-SEP-2000;  2000US-0231414.
PR   08-SEP-2000;  2000US-0232080.
PR   08-SEP-2000;  2000US-0232081.
PR   12-SEP-2000;  2000US-0231968.
PR   14-SEP-2000;  2000US-0232397.
PR   14-SEP-2000;  2000US-0232398.
PR   14-SEP-2000;  2000US-0232399.
PR   14-SEP-2000;  2000US-0232400.
PR   14-SEP-2000;  2000US-0232401.
PR   14-SEP-2000;  2000US-0233063.
PR   14-SEP-2000;  2000US-0233064.
PR   14-SEP-2000;  2000US-0233065.
PR   21-SEP-2000;  2000US-0234223.
PR   21-SEP-2000;  2000US-0234274.
PR   25-SEP-2000;  2000US-0234997.
PR   25-SEP-2000;  2000US-0234998.
PR   25-SEP-2000;  2000US-0235484.
PR   27-SEP-2000;  2000US-0235834.
PR   27-SEP-2000;  2000US-0235836.
PR   29-SEP-2000;  2000US-0236327.
PR   29-SEP-2000;  2000US-0236367.
PR   29-SEP-2000;  2000US-0236368.
PR   29-SEP-2000;  2000US-0236369.
PR   29-SEP-2000;  2000US-0236370.
PR   02-OCT-2000;  2000US-0236802.
PR   02-OCT-2000;  2000US-0237037.
PR   02-OCT-2000;  2000US-0237038.
PR   02-OCT-2000;  2000US-0237039.
PR   02-OCT-2000;  2000US-0237040.
PR   13-OCT-2000;  2000US-0239935.
PR   20-OCT-2000;  2000US-0239937.
PR   20-OCT-2000;  2000US-0240960.
PR   20-OCT-2000;  2000US-0241221.
PR   20-OCT-2000;  2000US-0241785.
PR   20-OCT-2000;  2000US-0241786.
PR   20-OCT-2000;  2000US-0241787.
PR   20-OCT-2000;  2000US-0241808.
PR   20-OCT-2000;  2000US-0241809.
PR   20-OCT-2000;  2000US-0241826.
PR   01-NOV-2000;  2000US-0244617.
PR   08-NOV-2000;  2000US-0244674.
PR   08-NOV-2000;  2000US-0244675.
PR   08-NOV-2000;  2000US-0244676.
PR   08-NOV-2000;  2000US-0244677.
PR   08-NOV-2000;  2000US-0244678.
PR   08-NOV-2000;  2000US-0246523.
PR   08-NOV-2000;  2000US-0246524.
PR   08-NOV-2000;  2000US-0246525.
PR   08-NOV-2000;  2000US-0246526.
PR   08-NOV-2000;  2000US-0246527.
PR   08-NOV-2000;  2000US-0246528.
PR   08-NOV-2000;  2000US-0246532.
PR   08-NOV-2000;  2000US-0246609.
PR   08-NOV-2000;  2000US-0246610.
PR   08-NOV-2000;  2000US-0246611.
PR   08-NOV-2000;  2000US-0246613.
PR   17-NOV-2000;  2000US-0249207.
PR   17-NOV-2000;  2000US-0249208.
PR   17-NOV-2000;  2000US-0249209.
PR   17-NOV-2000;  2000US-0249210.
PR   17-NOV-2000;  2000US-0249211.
PR   17-NOV-2000;  2000US-0249212.
PR   17-NOV-2000;  2000US-0249213.
PR   17-NOV-2000;  2000US-0249214.
PR   17-NOV-2000;  2000US-0249215.
PR   17-NOV-2000;  2000US-0249216.
PR   17-NOV-2000;  2000US-0249217.
PR   17-NOV-2000;  2000US-0249218.
PR   17-NOV-2000;  2000US-0249244.
PR   17-NOV-2000;  2000US-0249245.
PR   17-NOV-2000;  2000US-0249264.
```

PR 17-NOV-2000; 2000US-0249265.
PR 17-NOV-2000; 2000US-0249297.
PR 17-NOV-2000; 2000US-0249299.
PR 17-NOV-2000; 2000US-0249300.
PR 01-DEC-2000; 2000US-0250160.
PR 01-DEC-2000; 2000US-0250391.
PR 01-DEC-2000; 2000US-0250391.
PR 05-DEC-2000; 2000US-0251030.
PR 05-DEC-2000; 2000US-0251988.
PR 05-DEC-2000; 2000US-0256719.
PR 06-DEC-2000; 2000US-0251479.
PR 06-DEC-2000; 2000US-0251856.
PR 08-DEC-2000; 2000US-0251868.
PR 08-DEC-2000; 2000US-0251869.
PR 08-DEC-2000; 2000US-0251989.
PR 08-DEC-2000; 2000US-0251990.
PR 11-DEC-2000; 2000US-0254097.
PR 05-JAN-2001; 2001US-0259678.
XX
PA (HUMA-) HUMAN GENOME SCI INC.
XX
PI Rosen CA, Barash SC, Ruben SM;
XX
DR WPI; 2001-465570/50.
XX
PT Isolated nucleic acid molecule encoding a reproductive system antigen
PT is used in preventing, treating or ameliorating a medical condition -
XX
PS Disclosure; SEQ ID NO 7809; 1297pp + Sequence Listing; English.
XX
CC The present invention provides the protein and coding sequences of a
CC number of human reproductive system related antigens. These can be used
CC in the prevention and treatment of reproductive system disorders,
CC including cancer. The present sequence is a genomic sequence encoding a
CC protein of the invention.
XX
SQ Sequence 29163 BP; 6801 A; 7298 C; 7317 G; 7747 T; 0 other;

Query Match   20.9%; Score 209.4; DB 22; Length 29163;
Best Local Similarity   62.0%;
Matches 360; Conservative 0; Mismatches 201; Indels 20; Gaps 1;
Pred. No. 2.6e-42;

```
Qy      4 GCCACCATGCCTGGGCTAATTTTTCTTTATTTTTTAGTAGAGACGAGGTTTTGCCATGTTGACC 63
          |||||||||
Db   4179 GCCACCACGTGGCTTATTTTTGTTATTTTTTAGTAGAGACAGGGTGTCACCTTGTTGGCC 4120

Qy     64 AGGCTGGTCTCGAACTCTTGACCTCAGGTGATCCGCCTGCCTCAGCCTCCCAAAGTGCTG 123
          |||||||||| |||||
Db   4119 AGGCTGGTCTTGAACTCCGGACCTCAGCGTGATCCACCTGCCTCGGCCTCCCAAAGTGCTG 4060

Qy    124 GGATTATAGGCGTGAGCGCGCGCCACCCACGCCCAGCCAACATTTTTTAAATACTGAAAAGTAGAGG 183
          |||||||  ||||||||||||||||||||| |||||| ||
Db   4059 GAATTACAGGCGGTGAGCCACCACCGCCCGGCCTGTGTAATCCACCACTTTGGAGGCCGAGGC 4000

Qy    184 GAATAGTTATAGTGTACCCCATTTACCCATCACTCAGTTCAACAGCTGGTGACATATTT 243
          |||||||| | ||||| |||| |
Db   3999 AGATGGTTCACTTGCGCGCCCAGAGTTCCAGATCAGCCTGGGCAACATGGAGAAACTCTGT 3940

Qy    244 ATTTCTTCTATACCAGTACCGTACTCTCCCACTGGGATTATTTTTTAGGCAAAACCCAGA 303
          |||| |||| || ||| ||
Db   3939 CTGTACAAAAAATAAATTAGGCGGGCTGGGCACGTGG------------CTCA 3900

Qy    304 TGACATTTTATCCCTAAATACTTTAGTATAAAGGTGTTCTTTCGAAAAAAATCATAACCTCA 363
          ||| | |   |    |
Db   3899 TGCCTGTAATCCTAACACTTTGGGAGGTCGAGGCGGGCGGATCATGAAGTCAGGAGTTCG 3840

Qy    364 GGACCAGCCCTGGCCAACACATGGTGAAACCCTGTCTGTACTAAAAATACAAAAAATTAGCTG 423
          |||||||  |||||||||||||
Db   3839 AGACCAGCCTGACCAACACATGGTGAAACCCCATCTGTATTAAAAATATAAAAAATTAGCCAG 3780

Qy    424 GCATGGTGCGTGGGCACCTGTAATCCCAGGAAGTCAGGAGGAAGGAGGAAATCACTT 483
          ||| ||| ||
Db   3779 GCACGGTAGTGCACGCCTGTAATCCCAGCTACTCAGGAGGCTGAGGCAGGAGAATCTCTT 3720
```

RESULT 9
AB198013/c
ID   AB198013 standard; DNA; 29163 BP.
XX
AC   AB198013;
XX
DT   21-JUN-2002 (first entry)
XX
DE   Human testicular antigen encoding DNA fragment SEQ ID NO: 2665.
XX
KW   Human; testicular antigen; testes; cancer; metastasis; immune disorder;
KW   reproductive system disorder; urinary system disorder; gene therapy;
KW   cardiovascular disorder; respiratory disorder; neurological disorder;
KW   gastrointestinal disease; infection; cytostatic; gene; ds.
XX
OS   Homo sapiens.
XX
PN   WO200155317-A2.
XX
PD   02-AUG-2001.
XX
PF   17-JAN-2001; 2001WO-US01329.

```
Qy    484 GAATCCGGGAAGCAGAAGAGATTGCAGTGAGCTGAGGATTGCAGTCGAGCCTGGGCGCGGACAGAGA 543
          ||| ||||| | |||||  || ||||||  |||||||||||||  ||  |  |
Db   3719 GAACCCGGGGAGGTGGAGTTTGCAGTGAGCTGAGATCGACCACTGCACTCCACGCCTGGGC 3660

Qy    544 CAGAAATGAAACTCTGTCTCAAAAACAAAACAAAACAAAAAAAA 584
          ||||| || ||| || | ||||
Db   3659 GACAAAGCAGAGACTCCAGAAAAAAAAAAAAAAAAGAAAAGAAAA 3619
```

PR   31-JAN-2000; 2000US-0179065.
PR   04-FEB-2000; 2000US-0180628.
PR   24-FEB-2000; 2000US-0184664.
PR   26-JUL-2000; 2000US-0220963.
PR   26-JUL-2000; 2000US-0220964.
PR   02-MAR-2000; 2000US-0166350.
PR   16-MAR-2000; 2000US-0189874.
PR   17-MAR-2000; 2000US-0190076.
PR   18-APR-2000; 2000US-0188123.
PR   19-MAY-2000; 2000US-0205515.
PR   07-JUN-2000; 2000US-0209467.
PR   28-JUN-2000; 2000US-0214886.
PR   30-JUN-2000; 2000US-0215135.
PR   07-JUL-2000; 2000US-0216647.
PR   07-JUL-2000; 2000US-0216880.
PR   11-JUL-2000; 2000US-0217487.
PR   11-JUL-2000; 2000US-0217496.
PR   14-JUL-2000; 2000US-0218290.
PR   14-JUL-2000; 2000US-0218290.
PR   26-JUL-2000; 2000US-0220963.
PR   26-JUL-2000; 2000US-0220964.
PR   14-AUG-2000; 2000US-0224518.
PR   14-AUG-2000; 2000US-0224519.
PR   14-AUG-2000; 2000US-0225213.
PR   14-AUG-2000; 2000US-0225214.
PR   14-AUG-2000; 2000US-0225266.
PR   14-AUG-2000; 2000US-0225267.
PR   14-AUG-2000; 2000US-0225268.
PR   14-AUG-2000; 2000US-0225270.
PR   14-AUG-2000; 2000US-0254447.
PR   14-AUG-2000; 2000US-0255757.
PR   14-AUG-2000; 2000US-0255758.
PR   14-AUG-2000; 2000US-0255759.
PR   18-AUG-2000; 2000US-0226279.
PR   22-AUG-2000; 2000US-0226681.
PR   22-AUG-2000; 2000US-0226868.
PR   22-AUG-2000; 2000US-0227182.
PR   22-AUG-2000; 2000US-0227009.
PR   23-AUG-2000; 2000US-0227009.
PR   30-AUG-2000; 2000US-0228924.
PR   01-SEP-2000; 2000US-0229287.
PR   01-SEP-2000; 2000US-0229343.
PR   01-SEP-2000; 2000US-0229344.
PR   01-SEP-2000; 2000US-0229345.
PR   05-SEP-2000; 2000US-0229509.

```
PR   05-SEP-2000;  2000US-0225513.
PR   06-SEP-2000;  2000US-0230437.
PR   06-SEP-2000;  2000US-0230438.
PR   08-SEP-2000;  2000US-0231242.
PR   08-SEP-2000;  2000US-0231243.
PR   08-SEP-2000;  2000US-0231244.
PR   08-SEP-2000;  2000US-0231413.
PR   08-SEP-2000;  2000US-0231414.
PR   08-SEP-2000;  2000US-0232080.
PR   08-SEP-2000;  2000US-0232081.
PR   12-SEP-2000;  2000US-0231968.
PR   14-SEP-2000;  2000US-0232397.
PR   14-SEP-2000;  2000US-0232398.
PR   14-SEP-2000;  2000US-0232399.
PR   14-SEP-2000;  2000US-0232400.
PR   14-SEP-2000;  2000US-0232401.
PR   14-SEP-2000;  2000US-0233063.
PR   14-SEP-2000;  2000US-0233064.
PR   14-SEP-2000;  2000US-0233065.
PR   21-SEP-2000;  2000US-0234223.
PR   21-SEP-2000;  2000US-0234274.
PR   25-SEP-2000;  2000US-0234997.
PR   25-SEP-2000;  2000US-0234998.
PR   26-SEP-2000;  2000US-0235484.
PR   27-SEP-2000;  2000US-0235834.
PR   27-SEP-2000;  2000US-0235836.
PR   29-SEP-2000;  2000US-0236327.
PR   29-SEP-2000;  2000US-0236367.
PR   29-SEP-2000;  2000US-0236368.
PR   29-SEP-2000;  2000US-0236369.
PR   29-SEP-2000;  2000US-0236370.
PR   02-OCT-2000;  2000US-0236802.
PR   02-OCT-2000;  2000US-0237037.
PR   02-OCT-2000;  2000US-0237038.
PR   02-OCT-2000;  2000US-0237039.
PR   13-OCT-2000;  2000US-0239937.
PR   13-OCT-2000;  2000US-0239935.
PR   02-OCT-2000;  2000US-0237040.
PR   20-OCT-2000;  2000US-0409960.
PR   20-OCT-2000;  2000US-0241221.
PR   20-OCT-2000;  2000US-0241785.
PR   20-OCT-2000;  2000US-0241786.
PR   20-OCT-2000;  2000US-0241787.
PR   20-OCT-2000;  2000US-0241808.
PR   20-OCT-2000;  2000US-0241809.
PR   01-NOV-2000;  2000US-0244617.
PR   08-NOV-2000;  2000US-0246474.
PR   08-NOV-2000;  2000US-0246475.
PR   08-NOV-2000;  2000US-0246476.
PR   08-NOV-2000;  2000US-0246477.
PR   08-NOV-2000;  2000US-0246478.
PR   08-NOV-2000;  2000US-0246523.
PR   08-NOV-2000;  2000US-0246524.
PR   08-NOV-2000;  2000US-0246525.
PR   08-NOV-2000;  2000US-0246526.
PR   08-NOV-2000;  2000US-0246527.
PR   08-NOV-2000;  2000US-0246528.
PR   08-NOV-2000;  2000US-0246532.
PR   08-NOV-2000;  2000US-0246609.
PR   08-NOV-2000;  2000US-0246610.
PR   08-NOV-2000;  2000US-0246611.
PR   08-NOV-2000;  2000US-0246613.
PR   17-NOV-2000;  2000US-0249207.
PR   17-NOV-2000;  2000US-0249208.
PR   17-NOV-2000;  2000US-0249209.
PR   17-NOV-2000;  2000US-0249210.
PR   17-NOV-2000;  2000US-0249211.
PR   17-NOV-2000;  2000US-0249212.
PR   17-NOV-2000;  2000US-0249213.
PR   17-NOV-2000;  2000US-0249214.
PR   17-NOV-2000;  2000US-0249215.
PR   17-NOV-2000;  2000US-0249216.
PR   17-NOV-2000;  2000US-0249217.
PR   17-NOV-2000;  2000US-0249218.
PR   17-NOV-2000;  2000US-0249244.
PR   17-NOV-2000;  2000US-0249245.
PR   17-NOV-2000;  2000US-0249264.
PR   17-NOV-2000;  2000US-0249265.
PR   17-NOV-2000;  2000US-0249297.
PR   17-NOV-2000;  2000US-0249299.
PR   17-NOV-2000;  2000US-0249300.
PR   01-DEC-2000;  2000US-0250160.
PR   01-DEC-2000;  2000US-0250391.
PR   05-DEC-2000;  2000US-0251030.
PR   05-DEC-2000;  2000US-0251479.
PR   08-DEC-2000;  2000US-0251856.
PR   08-DEC-2000;  2000US-0251868.
PR   08-DEC-2000;  2000US-0251869.
PR   08-DEC-2000;  2000US-0251989.
PR   08-DEC-2000;  2000US-0251990.
PR   11-DEC-2000;  2000US-0254097.
PR   05-JAN-2001;  2001US-0259678.
XX
PA   (HUMA-) HUMAN GENOME SCI INC.
XX
PI   Rosen CA;  Barash SC,  Ruben SM;
XX
WPI; 2001-483232/52.
XX
DR
XX
PT   Nucleic acids encoding 973 human testicular antigen polypeptides,
PT   useful for preventing, diagnosing and/or treating testicular cancer -
XX
PS   Disclosure; SEQ ID NO 2665; 766pp; English.
XX
CC   The present invention provides the protein and coding sequences of 973
CC   human testicular antigens, and fragments of their genomic sequences. The
CC   sequences can be used in the treatment of cardiovascular, urinary system,
CC   reproductive system, immune, respiratory, neurological and
CC   gastrointestinal disorders, infections, and particularly cancer,
CC   especially testicular cancers. The present sequence is a DNA encoding a
CC   protein fragment of the invention.
XX
SQ   Sequence 29163 BP;  6801 A;  7298 C;  7317 G;  7747 T;  0 other;

Query Match     20.9%;  Score 209.4;  DB 23;  Length 29163;
Best Local Similarity   62.08%;  Pred. No. 2.6e-42;
Matches  360;  Conservative   0;  Mismatches  201;  Indels  20;  Gaps  1;

Qy      4 GCCACCATGCCTGGCTGAATTTCTTATTTTCTTATTTTAGTAGAGACGAGGTTTGCCATGTTGACC 63
          ||||||| | ||||||| ||||| |||||   ||||| ||| ||| ||| ||| |||||||||
Db   4179 GCCACCACGTGGCTTATTTTGTTATTTTAGTAGAGACAGGGTGTCACCTGTTGTTGGCC 4120

Qy     64 AGGCTGGTCTCGAAACTCTTGACCTCAGGTGATCGCCGCCTCCCAAAGTGCTG 123
          ||||||||||| ||||||  |||||||| |||| ||||| |||||||||||||
Db   4119 AGGCTGGTCTTGAAACTCCGGACCTCAGCTGATCCACCTGCCTGGCCTCCCAAAGTGCTG 4060

Qy    124 GGATTATAGGCGTGAGCCGCCGCCACCCGCCAGCCAACATTTTTTTAAATACTGAAAAGTAGAGG 183
          | ||||| |||| | |||| || |||||||||  |      ||||||     ||    |||| |||
Db   4059 GAATTACAGGCGGGTGAGCCACACCGCCACCCGCCACACTTTTGGAGGCCGAGGC 4000

Qy    184 GAATAGTTATATGTGTACCCCATTTACCCATCACTCAGTTTCAACAGCTGGTTGACATTATTT 243
          | ||  |||  ||  | |||    || |  |||| ||||  |||||| |  ||  |||| ||||
Db   3999 AGATGGTTCACTTGCGCCCAGAGTTCCAGATCAGCCTGGGCAACATGGAGAAACTCTGT 3940

Qy    244 ATTTCTTCTATACCAGTACCGTACTCTCCCCACTGGATTATTTTAAGGCAAAACCCAGA 303
          | ||  |  || |   |   |    |   || |   |  ||||| |  |  ||||||||
Db   3939 CTGTACAAAAAAATAAATTAGGCGGCTGGGCACAGTGG------ 3900

Qy    304 TGACATTTTATCCCTAAATACTTTAGATAAAGGTGTTCTTTGCAAAAAAATCATAACCTCA 363
          || | ||||| | |  |  || |||| ||  |  |||| |||| |||  |   ||  || ||
Db   3899 TGCCTGTAAATCCTAACACTTTGGGAGGTCGAGGGACGGATCATGAAGTCAGGAGTTCG 3840

Qy    364 GGACCAGCCTGGCCAACATGGTGAAACCCTGTCTGTACTAAAAATACAAAAATTAGCTTG 423
```

```
Db   3839 AGACCAGCCTGACCAACATGGTCGAAACCCATCTGTATTAAAATATAAAAATTAGCCAG  3780
          |||||||||           |||||    ||   |||||| |||||||| |||||||||
Qy    424 GCATGGTTCGTGGGCACCTGAATCCACCAGTACTCAGGAGACTGAGGGAGGAGAATCACTT  483
          ||| ||  ||||| ||  |||||||| |||  || |||||| |||||| |||||| ||| ||
Db   3779 GCACGGTAGTGCACGCTGTAATCCCAGCTACTCAGGAGGCTGAGGCAGGAGAATCTCTT  3720
Qy    484 GAATCCCGGAAGCAGCAGAGAGAGATTGCAGTGAGCTGAGATTGCGCCACTGCACTCCAGCCTGGGCGACAGAGA  543
          |||| || ||||| | ||  ||||||||||||||| |||||| | | |||| ||||||||||||| | ||| |
Db   3719 GAACCCGGGAAGCGGAGGTTGCAGTGAGCCGAGATCCGCCACCACTGCACTCCAGCCTGGGC  3660
Qy    544 CAGAAATGAAAACTCTGTCTCAAAAACAAAACAAAACAAAAAAAA  584
          ||| || |||| |||| |||  |||||||||||||||||| |||||
Db   3659 GACAAAGCAAGACTCCAGAAAAAAAAAAAAAAAAGAAAAAGAAAAA  3619
```

RESULT 10

ID   ABK83569 standard; cDNA; 122888 BP.
XX
AC   ABK83569;
XX
DT   14-AUG-2002 (first entry)
XX
DE   Human cDNA differentially expressed in granulocytic cells #140.
XX
KW   Human; ss; granulocytic cell; DNA chip; bacterial infection;
KW   viral infection; parasitic infection; protozoal infection;
KW   fungal infection; sterile inflammatory disease; psoriasis;
KW   rheumatoid arthritis; glomerulonephritis; asthma; thrombosis;
KW   cardiac reperfusion injury; renal reperfusion injury; ARDS;
KW   adult respiratory distress syndrome; inflammatory bowel disease;
KW   Crohn's disease; ulcerative colitis; periodontal disease;
KW   granulocyte activation; chronic inflammation; allergy.
XX
OS   Homo sapiens.
XX
CC   PN   WO20022899-A2.
CC   PD   11-APR-2002.
CC   PF   03-OCT-2001; 2001WO-US30821.
CC   PR   03-OCT-2000; 2000US-237189P.
CC   PA   (GENE-) GENE LOGIC INC.
CC   PI   Beazer-Barclay Y, Weissman SM, Yamaga S, Vockley J;
CC   DR   WPI; 2002-435328/46.
CC   PT   Detecting granulocyte activation by detecting differential expression
CC   PT   of genes associated with granulocyte activation, which serves as
CC   PT   diagnostic markers that is useful for monitoring disease states and
CC   PT   drug toxicity -
CC   PS   Claim 1; SEQ ID NO 140; 114pp; English.
CC   CC   The invention relates to detecting (M1) granulocyte (GC) activation
CC   (GCA), by detecting the level of expression of gene(s) (Gs) identified by
CC   DNA chip analysis as given in the specification, and comparing
CC   the expression level to an expression level in an unactivated
CC   GC, where differential expression of Gs is indicative of GCA.
CC   Also included are modulating (M2) GA by contacting Gc with an agent
CC   that alters the expression of at least one gene in Gs; (2) screening (M3)
CC   for an agent capable of modulating GCA or an inflammation (especially
CC   chronic) in a tissue, an allergic response in a subject, exposure of a
CC   subject to a pathogen or sterile inflammatory disease using the
CC   gene expression profile; (3) detecting (M4) an inflammation (especially
CC   chronic) in a tissue, an allergic response in a subject, exposure of a
CC   subject to a pathogen or sterile inflammatory disease, by detecting the
CC   level of expression in a sample of the tissue of gene(s) from Gs, where
CC   the level of expression of the gene is indicative of inflammation;
CC   CC   (4) treating (M5) an inflammation (especially chronic) or in a tissue,
CC   an allergic response in a subject, exposure of a subject to a pathogen
CC   or sterile inflammatory disease, by contacting the expression of gene(s)
CC   or inflammation with an agent that modulates the expression of gene(s)
CC   from Gs in the tissue. M1 is useful for detecting GCA; M2 is useful for
CC   modulating GA; M3 is useful for screening an agent capable of modulating
CC   GCA preferably an inflammation in a tissue; M4 is useful for
CC   detecting an inflammation (especially chronic) in a tissue, an allergic
CC   response in a subject, exposure of a subject to a pathogen or sterile
CC   inflammatory disease (e.g. psoriasis, rheumatoid arthritis,
CC   glomerulonephritis, asthma, thrombosis, cardiac reperfusion injury, renal
CC   reperfusion injury, ARDS, adult respiratory distress syndrome,
CC   inflammatory bowel disease, Crohn's disease, ulcerative colitis,
CC   periodontal disease; also bacterial infection, viral infection,
CC   parasitic infection, protozoal infection, fungal infection and M5 is
CC   useful for treating one of the above conditions. The present
CC   sequence represents a gene differentially expressed in granulocytes.
CC   Note: The sequence data for this patent did not form part
CC   of the printed specification, but was obtained in electronic
CC   format directly from WIPO at
CC   ftp.wipo.int/pub/published_pct_sequences.
XX
SQ   Sequence 122888 BP; 28761 A; 33410 C; 31919 G; 28798 T; 0 other;

Query Match   20.8%; Score 207.8; DB 24; Length 122888;
Best Local Similarity   63.0%; Pred. No. 1e-41;
Matches 371; Conservative   0; Mismatches 212; Indels   6; Gaps 3;

```
Qy      1 CCTGCCCACCATGCCTGGCTAATTTTTCTTTATTTTTAGTAGAGACGAGGGTTTGCCATGTTG   60
          |||||||||||       || || ||||||   |||| |||||||||| || |||||| |||||||
Db  38749 CCTGCCCACCACACCCGGTTAATTTTT-GTRTTTTTAGTAGAGACAGAGGGTTTCACCATGTTG  38691
Qy     61 ACCAGGCTGGTCTCGAACTCTTGACCTCAGGTGATCCGCCCGCCTCGGCCTCCCAAAGTG  120
          ||||| ||||||||||| || |||||||| |||||||||||||||||||||||||||||||
Db  38690 GCCAGACTGGTCTCGAAACTCCTGACCTCAAGTGATCCGCCCGCCTCGGCCTCCCAAAGTG  38631
Qy    121 CTGGGATTATAGGCGTGAGCCGCGCGCACCCAGCCAACATTTTTTAAATACTGAAAAGTAG  180
          |||||||||  ||||||||||  | ||  |||| | |||||     ||| |||  || |||
Db  38630 CTGGGATTACAGGCGTGAGCCTGCACCCGCCTTAACTACCAATAACTTACTGTTGAC  38571
Qy    181 AGGGAATAGTTATAGTGTACCCCATTTACCCTACCACTCAGTTTCAACAGCTGGTTGACATA  240
          | |||||| ||| ||| || ||| || || ||| || |||  || ||   |||| |  ||
Db  38570 TGTGAAAAACAGTTGATTAACACATATTTTGTGTGTTACCGTGTATTATATTAGTCTT  38511
Qy    241 TTTATTTCTTCTATACCAGTACCGTACTCTCCCCACTGGGATTATTTTTTAAGGCAAAACCC  300
          ||| ||| || ||| || || | |  ||   |  ||||||| ||||| |||| ||||| |||
Db  38510 ATAATAAAAGTAAGGTAGAGAAAAAGATGTTATTAAGATAATCTGGGCTGGGCATAGTGG  38451
Qy    301 AGATGACATTTTATCCCTAAATACTTAGATAAAGG---TGTTCTTTGAAAAAAATCATA  357
          |  | || || | || |||||| |||| | |||||   ||||| ||  |||||   ||
Db  38450 CTCATGCCTGTAATCTCAGCACTTTGGGAGGCCAAGGCAGGCGGACCCTCTGAGGTCAGG  38391
Qy    358 ACCTCAGGACCAGCCTGGCCAACATGGTGAAACCCCGTCTGTCTGTACTAAAAATACAAAATT  417
          |  | ||||||||||||||||||||||||||||||||||| | | || |||||||| ||||||||
Db  38390 AGTTCGAGACCAGCCTGGCCAACATGGTGAAACCCCGTCTCTACTAAAAATATAAAAATT  38331
Qy    418 AGCTTGGCATGGTGGCGTGTGGGCCACTGTAATCCCAGCTACTCAGGAGGCTGAGGCAGGAGA  477
          ||||  |||||||||||  | || |  |||| |||||||||||||||||||||||||| |||||
Db  38330 AGCTGGGCATGGTGGCGCGCGCCTGTTAGTCCCAGCTACTCAGGAGGCTGAGGCATGAGAA  38271
Qy    478 TCACTTGAATCCGGGAAGCAGAGATTGCAGTGAGCTGAGAT--TGCAGTCGAGCGCTGGGC  535
          || |||||| || |||| | ||| ||||||||||| |||||  |||| || ||| ||||||
Db  38270 TCGCTTGAACCCAGGAGGTGGAGGTTGCAGTGAGCCGAGATCCGCCATTGCACTCCAGC  38211
Qy    536 GACAGAGACAGAAATGAAAACTCTGTCTCAAAAACAAAACAAAACAAAAAAAA  584
          |||||| ||||| | ||  |||| | ||||||||||||||||||||||||||||||
Db  38210 CTGGGCGACAAGAGACTGAGACTCCATCTCAAAAAAAAAAAAAAAAAAAAAAGA  38162
```

RESULT 11

ID   AAK64975 standard; DNA; 946 BP.
XX
AC   AAK64975;
XX

```
AC   AAK64975;
XX
DT   06-NOV-2001  (first entry)
XX
DE   Human immune/haematopoietic antigen genomic sequence SEQ ID NO:19787.
XX
KW   Human; immune; haematopoietic; immune/haematopoietic antigen; cancer;
KW   cytostatic; gene therapy; vaccine; metastasis; ds.
XX
OS   Homo sapiens.
XX
PN   WO200157182-A2.
XX
PD   09-AUG-2001.
XX
PF   17-JAN-2001; 2001WO-US01354.
XX
PR   31-JAN-2000; 2000US-0179065.
PR   04-FEB-2000; 2000US-0180628.
PR   24-FEB-2000; 2000US-0184664.
PR   02-MAR-2000; 2000US-0186350.
PR   16-MAR-2000; 2000US-0189874.
PR   17-MAR-2000; 2000US-0190076.
PR   18-APR-2000; 2000US-0198123.
PR   19-MAY-2000; 2000US-0205515.
PR   07-JUN-2000; 2000US-0209467.
PR   28-JUN-2000; 2000US-0214886.
PR   30-JUN-2000; 2000US-0215135.
PR   07-JUL-2000; 2000US-0216647.
PR   07-JUL-2000; 2000US-0216880.
PR   11-JUL-2000; 2000US-0217487.
PR   11-JUL-2000; 2000US-0217496.
PR   14-JUL-2000; 2000US-0218290.
PR   26-JUL-2000; 2000US-0220963.
PR   26-JUL-2000; 2000US-0220964.
PR   14-AUG-2000; 2000US-0224518.
PR   14-AUG-2000; 2000US-0224519.
PR   14-AUG-2000; 2000US-0225213.
PR   14-AUG-2000; 2000US-0225214.
PR   14-AUG-2000; 2000US-0225266.
PR   14-AUG-2000; 2000US-0225267.
PR   14-AUG-2000; 2000US-0225268.
PR   14-AUG-2000; 2000US-0225270.
PR   14-AUG-2000; 2000US-0225447.
PR   14-AUG-2000; 2000US-0225757.
PR   14-AUG-2000; 2000US-0225758.
PR   14-AUG-2000; 2000US-0225759.
PR   18-AUG-2000; 2000US-0226279.
PR   22-AUG-2000; 2000US-0226681.
PR   22-AUG-2000; 2000US-0226868.
PR   22-AUG-2000; 2000US-0227182.
PR   23-AUG-2000; 2000US-0227009.
PR   30-AUG-2000; 2000US-0228924.
PR   01-SEP-2000; 2000US-0229287.
PR   01-SEP-2000; 2000US-0229343.
PR   01-SEP-2000; 2000US-0229344.
PR   01-SEP-2000; 2000US-0229345.
PR   05-SEP-2000; 2000US-0229509.
PR   05-SEP-2000; 2000US-0229513.
PR   06-SEP-2000; 2000US-0230437.
PR   06-SEP-2000; 2000US-0230438.
PR   08-SEP-2000; 2000US-0231242.
PR   08-SEP-2000; 2000US-0231243.
PR   08-SEP-2000; 2000US-0231244.
PR   08-SEP-2000; 2000US-0231413.
PR   08-SEP-2000; 2000US-0231414.
PR   08-SEP-2000; 2000US-0232080.
PR   08-SEP-2000; 2000US-0232081.
PR   12-SEP-2000; 2000US-0231968.
PR   14-SEP-2000; 2000US-0232397.
PR   14-SEP-2000; 2000US-0232398.
PR   14-SEP-2000; 2000US-0232399.
PR   14-SEP-2000; 2000US-0232400.

PR   14-SEP-2000; 2000US-0232401.
PR   14-SEP-2000; 2000US-0233063.
PR   14-SEP-2000; 2000US-0233064.
PR   14-SEP-2000; 2000US-0233065.
PR   21-SEP-2000; 2000US-0234223.
PR   21-SEP-2000; 2000US-0234274.
PR   25-SEP-2000; 2000US-0234997.
PR   25-SEP-2000; 2000US-0234998.
PR   26-SEP-2000; 2000US-0235484.
PR   27-SEP-2000; 2000US-0235834.
PR   27-SEP-2000; 2000US-0235835.
PR   29-SEP-2000; 2000US-0236327.
PR   29-SEP-2000; 2000US-0236367.
PR   29-SEP-2000; 2000US-0236368.
PR   29-SEP-2000; 2000US-0236369.
PR   29-SEP-2000; 2000US-0236370.
PR   02-OCT-2000; 2000US-0236802.
PR   02-OCT-2000; 2000US-0237037.
PR   02-OCT-2000; 2000US-0237038.
PR   02-OCT-2000; 2000US-0237039.
PR   02-OCT-2000; 2000US-0237040.
PR   13-OCT-2000; 2000US-0239935.
PR   13-OCT-2000; 2000US-0239937.
PR   20-OCT-2000; 2000US-0240960.
PR   20-OCT-2000; 2000US-0241221.
PR   20-OCT-2000; 2000US-0241785.
PR   20-OCT-2000; 2000US-0241786.
PR   20-OCT-2000; 2000US-0241787.
PR   20-OCT-2000; 2000US-0241808.
PR   20-OCT-2000; 2000US-0241809.
PR   20-OCT-2000; 2000US-0241826.
PR   01-NOV-2000; 2000US-0244617.
PR   08-NOV-2000; 2000US-0244474.
PR   08-NOV-2000; 2000US-0244475.
PR   08-NOV-2000; 2000US-0244476.
PR   08-NOV-2000; 2000US-0244477.
PR   08-NOV-2000; 2000US-0244478.
PR   08-NOV-2000; 2000US-0245523.
PR   08-NOV-2000; 2000US-0245524.
PR   08-NOV-2000; 2000US-0245525.
PR   08-NOV-2000; 2000US-0245526.
PR   08-NOV-2000; 2000US-0245527.
PR   08-NOV-2000; 2000US-0245528.
PR   08-NOV-2000; 2000US-0245532.
PR   08-NOV-2000; 2000US-0246509.
PR   08-NOV-2000; 2000US-0246610.
PR   08-NOV-2000; 2000US-0246611.
PR   17-NOV-2000; 2000US-0246513.
PR   17-NOV-2000; 2000US-0249207.
PR   17-NOV-2000; 2000US-0249208.
PR   17-NOV-2000; 2000US-0249209.
PR   17-NOV-2000; 2000US-0249210.
PR   17-NOV-2000; 2000US-0249211.
PR   17-NOV-2000; 2000US-0249213.
PR   17-NOV-2000; 2000US-0249214.
PR   17-NOV-2000; 2000US-0249215.
PR   17-NOV-2000; 2000US-0249216.
PR   17-NOV-2000; 2000US-0249217.
PR   17-NOV-2000; 2000US-0249218.
PR   17-NOV-2000; 2000US-0249244.
PR   17-NOV-2000; 2000US-0249245.
PR   17-NOV-2000; 2000US-0249264.
PR   17-NOV-2000; 2000US-0249265.
PR   17-NOV-2000; 2000US-0249297.
PR   17-NOV-2000; 2000US-0249299.
PR   17-NOV-2000; 2000US-0249300.
PR   01-DEC-2000; 2000US-0250160.
PR   01-DEC-2000; 2000US-0250391.
PR   05-DEC-2000; 2000US-0251030.
PR   05-DEC-2000; 2000US-0251988.
PR   05-DEC-2000; 2000US-0256719.
PR   06-DEC-2000; 2000US-0251479.
```

```
PR   08-DEC-2000; 2000US-0251856.
PR   08-DEC-2000; 2000US-0251868.
PR   08-DEC-2000; 2000US-0251869.
PR   08-DEC-2000; 2000US-0251990.
PR   08-DEC-2000; 2000US-0251990.
PR   11-DEC-2000; 2000US-0254097.
PR   05-JAN-2001; 2001US-0259678.
XX
PA   (HUMA-) HUMAN GENOME SCI INC.
XX
PI   Rosen CA, Barash SC, Ruben SM;
XX
DR   WPI; 2001-483426/52.
XX
PT   Nucleic acids encoding human immune/hematopoietic antigen polypeptides,
PT   useful for preventing, diagnosing and/or treating cancers and
PT   metastasis -
XX
PS   Disclosure; SEQ ID NO 19787; 3071pp + Sequence Listing; English.
XX
CC   AAK54951 to AAK64702 encode the human immune/haematopoietic antigen (I)
CC   amino acid sequences given in AAM82170 to AAM91921. (I) have cytostatic
CC   activity, and can be used in gene therapy and vaccine production. (I)
CC   proteins and polynucleotides may be used in the prevention, diagnosis and
CC   treatment of diseases associated with inappropriate (I) expression. For
CC   example, they may be used to treat disorders associated with decreased
CC   expression by rectifying mutations or deletions in a patient's genome
CC   that affect the activity of (I) by expressing inactive proteins or to
CC   supplement the patients own production of (I). Additionally, (I)
CC   polynucleotides may be used to produce the secreted (I), by inserting
CC   the nucleic acids into a host cell and culturing the cell to express the
CC   protein. (I) proteins and polynucleotides may be used to prevent,
CC   diagnose and treat immune/haematopoietic-related diseases, especially
CC   cancers and cancer metastases of haematopoietic-derived cells. AAK64703
CC   to AAK87694 represent human immune/haematopoietic antigen genomic
CC   sequences from the present invention. AAK54942 to AAK54950 and AAM82169
CC   represent sequences used in the exemplification of the present invention.
XX
SQ   Sequence 946 BP; 228 A; 232 C; 200 G; 286 T; 0 other;

Query Match      20.6%;   Score 206.2;   DB 22;   Length 946;
Best Local Similarity   61.08%;   Pred. No. 5.6e-42;
Matches 370;   Conservative 0;   Mismatches 233;   Indels 4;   Gaps 2;

Qy     4 GCCACCATGCCTGGCTRAATTTCTTRAATTTTNAGTRAGAGACGAGGTTTTGCCATGTTGACC 63
            ||||||||||||||| | |||||||||||||   |||||||||| ||||||||||||||||
Db   893 GCCATCATGCCTGGCTRAATTTTTTGTRATTTTTTGTRAGAGATGGGATTTTGCCATATTGCCC 834

Qy    64 AGGCTGGTCTCGAACTCTTGACCTCAGGTGRATCGGCCGCCTCCAGCCTCCCAAAGTGCTG 123
            |||||||| |  |||||||| |||||||||||||  ||||||||||||||||||||||||
Db   833 AGGCTGGTCTCCCAACTCTTGGGCTCAAGTGRATTCACCCGCCTCAGCCTCCCAAAGTGTTG 774

Qy   124 GGATTATAGGCGTGAGCGCGCGCCCACCCAGCCCAACA-TTTTTTAAATACTGAAAAGTAGA 181
            |||  |||  |||||||||||  ||| |||||||||  ||||||  || | |||||| |||
Db   773 GGACTACAGGGCGTGAGCCACCGCGCCCCAAGAAGATATRCAGATTTTGGTCAACAAG 714

Qy   182 GGGAATAGTTATAGTGTTACCCCCATTRACACTCAGTTTCACACAGCTGGGTGACATAT 241
            |||| ||| ||| ||||||||||||||||||||||||||||||||||||||||||| |
Db   713 CGCTGAAAAGAATGCTCAACATCATCAGTCATTTAGGGAAACTGCAAATCAAAAACACACACA 654

Qy   242 TTATTTCTTCTATACCAGTACCGTACTCTCCCACTGGGATTATTTTAAGGCAAAACCCA 301
            ||||  ||||| ||||||||||||||||||||||||||||||||||||||||||||||
Db   653 GATACTCTTTCACACCACTAGGATGGCTATAATTAAAACAACAGGGTCGGGGTGGTGGCT 594

Qy   302 GATGACANTTTTATCCCTAAAATACTTTAGATAAAG--GTGTTCTTTGAAAAAAATCATAAC 359
            | |||| || | | ||  |||  ||||| |||||   |||| || | ||||||||  |||
Db   593 TACGCCTGTTATCCTAGCACTTTGGGAGGCTGAGGCGGGTGATCACTTGAGGTCAGGAG 534

Qy   360 CTCAGGACCAGCCTGGCCAACATGGTGAAAACCCTGTCTGTACTAAAAAATACAAAAATTAG 419
            ||||||||| |||||||||||||||||||  |||||||| || |||||||| |||||||||
Db   533 TTTGAGACCACCCTGGCCAACATGGTGAAACCCTGTCTCTAATAAAAAATGCAAAAATTAG 474

Qy   420 CTTGGCATGGTGGTCGTGGGCACCTGTAATCCCAGCTACTCAGGAGGCTGAGGCAGGAGATC 479
            |  | |||||||  ||||| || || | ||  ||||| |||| ||||||||||||| ||||
Db   473 CCGGGCATGGTGGCAGGCAGCACCTGTRAATCCCAGCTACTTGGGAGGCTGAGGCAGGAGATC 414

Qy   480 ACTTGAATCCGGGAAGCAGCAGAGATTGCAGTGAGCTGAGATTGCAGTCGAGCCTGGGCGACA 539
            |||||||||| |||| | |||||||||||||||||||||||||||||||||||  ||||
Db   413 GCTTGAACCCAGGAGGCAGAGGTTGCTGGTGAGCCAAGATTGCACCATTGCACTCCAGCCT 354

Qy   540 GAGACAGAAATGAAACTCTGTTCTCAAAAACAAAACACAAAACCACTATACATAAA 599
            ||||||||  | | || |  | |  || ||||||||||||||||||| || ||
Db   353 GGGTGACAGAGTGAGACTTTCTCTCAAAAATAAATAAGATAAAATAAAATAAAAAGA 294

Qy   600 ATGAACA 606
            | |||
Db   293 AGCAACA 287
```

RESULT 12
AAS33462
```
ID   AAS33462 standard; DNA; 25715 BP.
XX
AC   AAS33462;
XX
DT   04-DEC-2001 (first entry)
XX
DE   DNA encoding human secreted protein, Seq ID NO 745.
XX
KW   Immunomodulatory; human immunodeficiency virus; HIV; anaemia; angina;
KW   rheumatoid arthritis; antiarteriosclerotic; cardiant; vascular;
KW   cerebroprotective; thrombolytic; antimicrobial; ophthalmological;
KW   cytostatic; Alzheimer's disease; Parkinson's disease; human; cancer;
KW   multiple sclerosis; cancer; hyperproliferative disorder; infection;
KW   Gaucher's disease; neurological disease; cerebrovascular disorder;
KW   thrombosis; wound healing; ds.
XX
OS   Homo sapiens.
XX
PN   WO200015326-A2.
XX
PD   02-AUG-2001.
XX
PF   17-JAN-2001; 2001WO-US01347.
XX
PR   31-JAN-2000; 2000US-0179065.
XX
PA   (HUMA-) HUMAN GENOME SCI INC.
XX
PI   Rosen CA, Barash SC, Ruben SM;
XX
DR   WPI; 2001-451931/48.
XX
PT   New nucleic acids and polypeptides, useful for diagnosing, preventing
PT   or treating medical conditions -
XX
PS   Disclosure; SEQ ID No 745; 753pp; English.
XX
CC   The invention relates to novel isolated nucleic acid molecules (I)
CC   encoding human secreted proteins (II). (I) and (II) are used to prevent,
CC   treat or ameliorate a medical condition in e.g. humans, mice, rabbits,
CC   goats, horses, cats, dogs, chickens or sheep. (I) and (II) may be used in
CC   the prevention, treatment and diagnosis of diseases associated with
CC   inappropriate expression of secreted proteins. (I) and complementary
CC   sequences may also be used as DNA probes in diagnostic assays (e.g.
CC   polymerase chain reactions (PCR)) to detect and quantitate the presence
CC   of similar nucleic acid sequences in samples, and so which patients may
CC   be in need of restorative therapy. (II) may also be used as antigens in
CC   the production of antibodies and in assays to identify modulators
CC   (agonists and antagonists) of the expression and activity of the secreted
CC   proteins. The anti-(II) antibodies and antagonists may also be used to
CC   down regulate expression and activity of (II). The anti-(II) antibodies
CC   may also be used as diagnostic agents for detecting the presence of (II)
CC   in samples (e.g. by enzyme linked immunosorbent assay (ELISA)). The
CC   disorders include for example: immune/autoimmune diseases (e.g. HIV
CC   (human immunodeficiency virus) infections, anaemia, rheumatoid arthritis
```

```
CC  and multiple sclerosis), cancers and hyperproliferative disorders (e.g.
CC  melanomas, neoplasms of the breast or liver, Sezary syndrome and
CC  Gaucher's disease), neurological diseases (e.g. Alzheimer's disease,
CC  Parkinson's disease and Charcot-Marie-Tooth disease), cardio-/
CC  cerebrovascular disorders (e.g. cardiac arrest, tachycardia, cardio-/
CC  angina and thrombosis), infections caused by bacteria, viruses and
CC  fungi and ocular disorders (e.g. corneal infections). (I) and (II),
CC  agonists, antagonists and antibodies can also be used to promote wound
CC  healing, maintain organs before transplantation, and support cell culture
CC  of primary tissues. AAS33043-AAS33486 represent human secreted protein
CC  coding sequences, PCR primers, and related sequences of the invention.
CC  Note: The sequence data for this patent did not appear in the printed
CC  specification but was obtained in electronic format directly from WIPO
CC  at: ftp.wipo.int/pub/published_pct_sequences.
XX
SQ  Sequence 25715 BP; 6112 A; 6093 C; 6275 G; 7177 T; 58 other;

Query Match         20.6%;  Score 205.8;  DB 22;  Length 25715;
Best Local Similarity   62.2%;  Pred. No. 2e-41;
Matches  374;  Conservative   0;  Mismatches 222;  Indels   5;  Gaps  3;

Qy      1 CCTGCCACCATGCCTGGGTAATTTTCTTTATTTTTTAGTAGAGACGGGGTTTTGCCATGTTG 60
          ||  ||| | |||  ||| ||||||| |  |||| |||||||||||| |||||||| ||||||
Db   8587 CCCACGACGACGGCCTGGCCTAATTTTTTGTTGTAATTTTTAGTAGAGATGGGGTTTCATGTTG 8646

Qy     61 ACCAGGCTGGTCTCGAACTCTTGACCTCCAGGTGATCGCCCGTCGCCTCCCAAAGTG 120
          | |||||||||||| ||||| |||||||||||||||||  |||| || |||||||||||||||
Db   8647 GCCACGCTGGTCTTGAACTCCTGACCTCCAGGTGATCTCGCCCATCTTGGCCTCCCAAAGTG 8706

Qy    121 CTGGGATTATAGGCGTGAGCCGCCGCGCCCAGCCAACATTTTTTAAAATACTGAAAAGTAG 180
          |||||||||| | ||||||||  ||||||||||||||  | |  || |||| ||| ||||
Db   8707 CTGGGATTACAGATGTGAGCCACTGTGCCCAGCCAATAATGTGCCTTTCATTTGA---TCC 8763

Qy    181 AGGGAATAGTTATTAGTGTGTACCCCATTTACCCATCACTCAGTTTCAACAGCTGGTGACATA 240
          | |||  || | |   |   |||  |||  || |||||||||| |||||| |||||||| ||
Db   8764 TGAAACCCCTTGAGGTAGTAGGATTGGCAGTTATCTCCATTTGATCAAGGACAAATAGAC 8823

Qy    241 TTTTATTTCTTCTATACCAGTACCGTACTCTCCCCACTGGGATTATTTTTTAAGGCAAAACCC 300
          | | |  | ||  |||| ||  | |||  |   ||| | ||||  || |||  ||| |||||
Db   8824 TAAAGCACAGAGAGAGAAAATAAATTTGCTTAAGATCATATAGTAGGCCGGGCGTGGTGGC 8883

Qy    301 AGATGACACATTTTATCCCTAAAATACTTTAGATAAAGGTGTTCTTTTGAAAAAAAATCATAACC 360
          | |||| ||  | | |ccc|     |   || |    | | |||    || |||| || ||||
Db   8884 TCATGCCCTGTAATCCCAGCACTTTGGGAGGCCGAGGCGGGATCACAAGGTCAGGAGA 8943

Qy    361 TCAGGACCAGCCTGGCCAACACATGGTGAAACCCTGTCTGTACTAAAAATAC-AAAAATTAG 419
          |||||||||||||||||||||||| |||||||||| ||| |||||||||||| |||| ||||
Db   8944 TCGAGACCATTCTGGCTAACACGGTGAAACCCCAGTCTCTACTAAAAATACAAAAAATTAG 9003

Qy    420 CTTGGCCATGGTCGTCGTGGGCACCTGTAATCCCAGCTACTCAGGAGGCTGAGGCAGGAGAATC 479
          |  ||| ||||| | ||ggg|  ||||| ||||||||||||||||||||||||||||||||||
Db   9004 CTGGGCGTGGTGGCGGGCGTTGCCTGTGGTCCCAGCTACTCAGGAGGCTGAGGCAGGAGAATG 9063

Qy    480 ACTTGAATCCGGGAAGCAAGAGATTGCAGTGAGCTGAGATTGCAGTCGAGCCTGGGCGGACA 539
          |   | ||||||||| | | | |||||||||||| ||||  |   ||| ||||||||| ||
Db   9064 GCATAAACCCGGGAGGCGGAGGTTGCAGTGAGCAGAGATCGTGCCACTGCACTCCAGCCT 9123

Qy    540 GAGACAGAAATGAA-ACTCTGTCTCAAAAACAAACAAACAAAAAAACCACTATACATATAAA 598
          | | | | |  ||  |||| |||||||||| |||     |   |||||   ||| || |||
Db   9124 GGGCGACAGAGCAAGACTCCGTCTCAAAAAAAAAAGATCAATATTAGTAA 9183

Qy    599 A 599
          |
Db   9184 A 9184
```

```
RESULT 13
AAK82338
ID  AAK82338 standard; DNA; 48908 BP.
XX
AC  AAK82338;
XX
DT  07-NOV-2001 (first entry)
```

```
PR   14-SEP-2000; 2000US-0233065.
PR   21-SEP-2000; 2000US-0234223.
PR   21-SEP-2000; 2000US-0234274.
PR   25-SEP-2000; 2000US-0234997.
PR   25-SEP-2000; 2000US-0234998.
PR   26-SEP-2000; 2000US-0235484.
PR   27-SEP-2000; 2000US-0235836.
PR   27-SEP-2000; 2000US-0235837.
PR   29-SEP-2000; 2000US-0236327.
PR   29-SEP-2000; 2000US-0236367.
PR   29-SEP-2000; 2000US-0236368.
PR   29-SEP-2000; 2000US-0236369.
PR   29-SEP-2000; 2000US-0236370.
PR   02-OCT-2000; 2000US-0236802.
PR   02-OCT-2000; 2000US-0237037.
PR   02-OCT-2000; 2000US-0237038.
PR   02-OCT-2000; 2000US-0237039.
PR   02-OCT-2000; 2000US-0237040.
PR   13-OCT-2000; 2000US-0239935.
PR   13-OCT-2000; 2000US-0239937.
PR   20-OCT-2000; 2000US-0240960.
PR   20-OCT-2000; 2000US-0241221.
PR   20-OCT-2000; 2000US-0241785.
PR   20-OCT-2000; 2000US-0241786.
PR   20-OCT-2000; 2000US-0241787.
PR   20-OCT-2000; 2000US-0241808.
PR   20-OCT-2000; 2000US-0241809.
PR   20-OCT-2000; 2000US-0241826.
PR   01-NOV-2000; 2000US-0244617.
PR   08-NOV-2000; 2000US-0246474.
PR   08-NOV-2000; 2000US-0246475.
PR   08-NOV-2000; 2000US-0246476.
PR   08-NOV-2000; 2000US-0246477.
PR   08-NOV-2000; 2000US-0246478.
PR   08-NOV-2000; 2000US-0246523.
PR   08-NOV-2000; 2000US-0246524.
PR   08-NOV-2000; 2000US-0246525.
PR   08-NOV-2000; 2000US-0246526.
PR   08-NOV-2000; 2000US-0246527.
PR   08-NOV-2000; 2000US-0246528.
PR   08-NOV-2000; 2000US-0246532.
PR   08-NOV-2000; 2000US-0246609.
PR   08-NOV-2000; 2000US-0246610.
PR   08-NOV-2000; 2000US-0246611.
PR   08-NOV-2000; 2000US-0246613.
PR   17-NOV-2000; 2000US-0249207.
PR   17-NOV-2000; 2000US-0249208.
PR   17-NOV-2000; 2000US-0249209.
PR   17-NOV-2000; 2000US-0249210.
PR   17-NOV-2000; 2000US-0249211.
PR   17-NOV-2000; 2000US-0249212.
PR   17-NOV-2000; 2000US-0249213.
PR   17-NOV-2000; 2000US-0249214.
PR   17-NOV-2000; 2000US-0249215.
PR   17-NOV-2000; 2000US-0249216.
PR   17-NOV-2000; 2000US-0249217.
PR   17-NOV-2000; 2000US-0249218.
PR   17-NOV-2000; 2000US-0249244.
PR   17-NOV-2000; 2000US-0249245.
PR   17-NOV-2000; 2000US-0249264.
PR   17-NOV-2000; 2000US-0249265.
PR   17-NOV-2000; 2000US-0249297.
PR   17-NOV-2000; 2000US-0249299.
PR   17-NOV-2000; 2000US-0249300.
PR   01-DEC-2000; 2000US-0250160.
PR   01-DEC-2000; 2000US-0250391.
PR   05-DEC-2000; 2000US-0251030.
PR   05-DEC-2000; 2000US-0251479.
PR   05-DEC-2000; 2000US-0251988.
PR   06-DEC-2000; 2000US-0256719.
PR   06-DEC-2000; 2000US-0251856.
PR   08-DEC-2000; 2000US-0251868.
PR   08-DEC-2000; 2000US-0251869.
```

```
PR   08-DEC-2000; 2000US-0251989.
PR   08-DEC-2000; 2000US-0251990.
PR   11-DEC-2000; 2000US-0254097.
PR   05-JAN-2001; 2001US-0259678.
XX
PA   (HUMA-) HUMAN GENOME SCI INC.
XX
PI   Rosen CA,  Barash SC,  Ruben SM;
XX
DR   WPI; 2001-483426/52.
XX
PS   Disclosure; SEQ ID NO 37150; 307lpp + Sequence Listing; English.
XX
CC   AAK54951 to AAK64702 encode the human immune/haematopoietic antigen (I)
CC   amino acid sequences given in AAM82170 to AAM91921. (I) have cytostatic
CC   activity, and can be used in gene therapy and vaccine production. (I)
CC   proteins and polynucleotides may be used in the prevention, diagnosis and
CC   treatment of diseases associated with decreased expression. For
CC   example, they may be used to treat disorders associated with decreased
CC   expression by rectifying mutations or deletions in a patient's genome
CC   that affect the activity of (I) by expressing inactive proteins or to
CC   supplement the patients own production of (I). Additionally, (I)
CC   polynucleotides may be used to produce the secreted (I), by inserting
CC   the nucleic acids into a host cell and culturing the cell to express the
CC   protein. (I) proteins and polynucleotides may be used to prevent,
CC   diagnose and treat immune/haematopoietic-related diseases, especially
CC   cancers and cancer metastases of haematopoietic-derived cells. AAK64703
CC   to AAK87694 represent human immune/haematopoietic antigen genomic
CC   sequences from the present invention. AAK54942 to AAK54950 and AAM82169
CC   represent sequences used in the exemplification of the present invention.
XX
SQ   Sequence 48908 BP; 14764 A; 10631 C; 10740 G; 12773 T; 0 other;

Query Match             20.6%;  Score 205.8;  DB 22;  Length 48908;
Best Local Similarity   61.6%;  Pred. No. 2.4e-41;
Matches  363;  Conservative   0;  Mismatches 222;  Indels   4;  Gaps  2;

Qy         4 GCCACCATGGCCTGGCTAATTTCTTATTTTTTTAGTAGAGACGAGGTTTGCCATGTTGACC 63
             ||||||| ||||||||| |||| |||||| |||||| |||||  |||||||||||| ||||
Db     46806 GCCACCACGCCTGGCAAATTTTGTATTTTTTAGAGACAGAGGTTTCTCCATGTTGCC 46865

Qy        64 AGGCTGGTCGTCGAACTCTTGACCTCAGGTGAATCCGCCTGCCTCGGCCTCCCAAAGTGCTG 123
             ||||||||||||||| |||| ||||||||| |||||| ||||||||||||||||||| ||||
Db     46866 AGGCTGGTCGTCGAACTCCCGACCTCAGGTGATCTGCCCACCTCGGCCTCCCAAAAATGCTG 46925

Qy       124 GGATTATAGGCGTGAGCCGCGCACCCAG--CCAACATATTTTTAAATACTGAAAAGTAGA 181
             ||||| |||  || ||||| ||| |||    |||| | |  ||||||||||||| ||||||
Db     46926 GGATTACAGGCATGAGCCACCACGTGCCCTGCCTCCAAAGGATCAGTTTAACAAACACAACAG 46985

Qy       182 GGGAATAGTTATAGTGTACCCCATTTACCCCATCACTCAGTTTCAAACAGCTGGTTGACATAT 241
             ||||||||||||| |||| |||||||||||||||||||| |||| |||| |||| ||| |||
Db     46986 ATGATTTGAAGCAAAACAGTACTCCAGTCACACTTTGACTACCAAAATGTTCATCCCTGCA 47045

Qy       242 TTATTTCTTCTATACCAGTACCCGTACTCTCCCACTGGGATTATTTTAAGGCAAAACCA 301
             |||| ||||| |||  |    |||
Db     47046 CCATCTCCTCTGTGGGGAGTCGTGAGAAGAAGAAATAACCGGAAGGTGCCAGGCGGTGGCT 47105

Qy       302 GATGACATTTTATCCCTAAATACTTTAGATAAAG--GTGTTCTTTGAAAAAAATCATAAC 359
             |||  | |        |  || |||||||||  |    |||| | ||||| ||||||| |||
Db     47106 CACGCCCTGTAATCCCAGCACTTTGGGAGGCCAAGGCGGGTAAATCACCTCAGGTGTCAGGAG 47165

Qy       360 CTCAGGACCAGCCTGGCCAACATGGTGAAACCCTGTCTGTTACTAAAAATACAAAAATTAG 419
             ||||||||||||||||||||||||||||||||||| ||| |||||||||| |||||||||||
Db     47166 TTCAAGACCAGCCTGGCCAACAGTGGTGAAACCCGCGTCTCTATTAAAAATACAAAAATTAG 47225

Qy       420 CTTGGCATGGTGGCGGGCGCAACCTGTAATCCCAGCTACTCAGGAGAACGTGAGGCAGGAGAATC 479
             || || |||||||| ||| ||| |||||||||||||||||||||||||||||||||||||||||
Db     47226 CCAGGTGAGGCGGGCGGTGGGGCGCCTGTAATCCCAGCTACTCGGGAGGCTGAGGCAGGAGAATT 47285
```

```
Qy    480  ACTTGAATCCGGGAAGCACAGAGATTGCAGTTGCAGTTGCAGTCGAGCCTGGGCGACA  539
            ||||||  |  ||||| || |||||  ||||||||  ||  |              |
Db  47286  GCTTGAACCTGGGAGGCCGGAGGTTGCAGAGAGTCGAGATCGCATCATTGCTCTCCAGCCT  47345

Qy    540  GAGACAGAAAATGAAACTCTGTCTCTCAAAAACAAACAAACAAAAAAAACCAC  588
            | ||||  |  ||  ||||||||  |  ||||  || ||| ||||||  ||
Db  47346  GGGTGACATGAGCAAAACTGTGTCTCAAAAAAAAAAAAAAAAAAAAATAC  47394
```

RESULT 14
ABA15328
ID   ABA15328 standard; DNA; 32192 BP.
XX
AC   ABA15328;
XX
DT   23-JAN-2002 (first entry)
XX
DE   Human nervous system related polynucleotide SEQ ID NO 7659.
XX
KW   Human; nootropic; neuroprotective; cytostatic; dermatological; virucide;
KW   immunosuppressive; antiinflammatory; anti-HIV; antibacterial; vulnerary;
KW   antiparkinsonian; antisickling; antianaemic; antiarthritic; cancer;
KW   antirheumatic; hepatotropic; cerebroprotective; antiinflammatory;
KW   antiallergic; antidiabetic; antiulcer; anticonvulsant; antifungal;
KW   antiparasitic; cardiant; immune disorder; cardiovascular disorder;
KW   neurological disease; infection; nephrotropic; gene therapy; vaccine; ds.
XX
OS   Homo sapiens.
XX
PN   WO20015906 3-A2.
XX
PD   16-AUG-2001.
XX
PF   17-JAN-2001; 2001WO-US01334.
XX
PR   31-JAN-2000; 2000US-0179065.
PR   04-FEB-2000; 2000US-0180628.
PR   24-FEB-2000; 2000US-0184664.
PR   02-MAR-2000; 2000US-0186350.
PR   16-MAR-2000; 2000US-0189874.
PR   17-MAR-2000; 2000US-0190076.
PR   18-APR-2000; 2000US-0198123.
PR   19-MAY-2000; 2000US-0205515.
PR   07-JUN-2000; 2000US-0209467.
PR   28-JUN-2000; 2000US-0214886.
PR   30-JUN-2000; 2000US-0215135.
PR   07-JUL-2000; 2000US-0216647.
PR   07-JUL-2000; 2000US-0216880.
PR   11-JUL-2000; 2000US-0217487.
PR   14-JUL-2000; 2000US-0217496.
PR   26-JUL-2000; 2000US-0220963.
PR   26-JUL-2000; 2000US-0220964.
PR   14-AUG-2000; 2000US-0224518.
PR   14-AUG-2000; 2000US-0224519.
PR   14-AUG-2000; 2000US-0225213.
PR   14-AUG-2000; 2000US-0225214.
PR   14-AUG-2000; 2000US-0225266.
PR   14-AUG-2000; 2000US-0225267.
PR   14-AUG-2000; 2000US-0225268.
PR   14-AUG-2000; 2000US-0225270.
PR   14-AUG-2000; 2000US-0225447.
PR   14-AUG-2000; 2000US-0225757.
PR   14-AUG-2000; 2000US-0225758.
PR   18-AUG-2000; 2000US-0225759.
PR   14-AUG-2000; 2000US-0226279.
PR   22-AUG-2000; 2000US-0226681.
PR   22-AUG-2000; 2000US-0226868.
PR   22-AUG-2000; 2000US-0227182.
PR   23-AUG-2000; 2000US-0227009.
PR   30-AUG-2000; 2000US-0228924.
PR   01-SEP-2000; 2000US-0229287.
PR   01-SEP-2000; 2000US-0229343.
```

```
PR    01-SEP-2000;  2000US-0229344.
PR    01-SEP-2000;  2000US-0229345.
PR    05-SEP-2000;  2000US-0229509.
PR    05-SEP-2000;  2000US-0229513.
PR    06-SEP-2000;  2000US-0230437.
PR    06-SEP-2000;  2000US-0230438.
PR    08-SEP-2000;  2000US-0231242.
PR    08-SEP-2000;  2000US-0231243.
PR    08-SEP-2000;  2000US-0231244.
PR    08-SEP-2000;  2000US-0231413.
PR    08-SEP-2000;  2000US-0231414.
PR    08-SEP-2000;  2000US-0232080.
PR    08-SEP-2000;  2000US-0232081.
PR    12-SEP-2000;  2000US-0231968.
PR    14-SEP-2000;  2000US-0232397.
PR    14-SEP-2000;  2000US-0232398.
PR    14-SEP-2000;  2000US-0232399.
PR    14-SEP-2000;  2000US-0232401.
PR    14-SEP-2000;  2000US-0233061.
PR    14-SEP-2000;  2000US-0233063.
PR    14-SEP-2000;  2000US-0233064.
PR    14-SEP-2000;  2000US-0233065.
PR    21-SEP-2000;  2000US-0234223.
PR    21-SEP-2000;  2000US-0234274.
PR    25-SEP-2000;  2000US-0234997.
PR    25-SEP-2000;  2000US-0234998.
PR    26-SEP-2000;  2000US-0235484.
PR    27-SEP-2000;  2000US-0235834.
PR    29-SEP-2000;  2000US-0235836.
PR    29-SEP-2000;  2000US-0236327.
PR    29-SEP-2000;  2000US-0236367.
PR    29-SEP-2000;  2000US-0236368.
PR    29-SEP-2000;  2000US-0236369.
PR    29-SEP-2000;  2000US-0236370.
PR    02-OCT-2000;  2000US-0236802.
PR    02-OCT-2000;  2000US-0237037.
PR    02-OCT-2000;  2000US-0237038.
PR    02-OCT-2000;  2000US-0237039.
PR    02-OCT-2000;  2000US-0237040.
PR    13-OCT-2000;  2000US-0239935.
PR    13-OCT-2000;  2000US-0239937.
PR    20-OCT-2000;  2000US-0240960.
PR    20-OCT-2000;  2000US-0241785.
PR    20-OCT-2000;  2000US-0241786.
PR    20-OCT-2000;  2000US-0241787.
PR    20-OCT-2000;  2000US-0241808.
PR    20-OCT-2000;  2000US-0241809.
PR    20-OCT-2000;  2000US-0241826.
PR    01-NOV-2000;  2000US-0242221.
PR    08-NOV-2000;  2000US-0244517.
PR    08-NOV-2000;  2000US-0246474.
PR    08-NOV-2000;  2000US-0246476.
PR    08-NOV-2000;  2000US-0246477.
PR    08-NOV-2000;  2000US-0246478.
PR    08-NOV-2000;  2000US-0246523.
PR    08-NOV-2000;  2000US-0246524.
PR    08-NOV-2000;  2000US-0246525.
PR    08-NOV-2000;  2000US-0246526.
PR    08-NOV-2000;  2000US-0246527.
PR    08-NOV-2000;  2000US-0246528.
PR    08-NOV-2000;  2000US-0246532.
PR    08-NOV-2000;  2000US-0246609.
PR    08-NOV-2000;  2000US-0246610.
PR    08-NOV-2000;  2000US-0246611.
PR    08-NOV-2000;  2000US-0466613.
PR    17-NOV-2000;  2000US-0249207.
PR    17-NOV-2000;  2000US-0249208.
PR    17-NOV-2000;  2000US-0249209.
PR    17-NOV-2000;  2000US-0249210.
PR    17-NOV-2000;  2000US-0249211.
PR    17-NOV-2000;  2000US-0249212.
PR    17-NOV-2000;  2000US-0249213.
```

PR   17-NOV-2000; 2000US-0249214.
PR   17-NOV-2000; 2000US-0249215.
PR   17-NOV-2000; 2000US-0249216.
PR   17-NOV-2000; 2000US-0249217.
PR   17-NOV-2000; 2000US-0249218.
PR   17-NOV-2000; 2000US-0249244.
PR   17-NOV-2000; 2000US-0249245.
PR   17-NOV-2000; 2000US-0249264.
PR   17-NOV-2000; 2000US-0249265.
PR   17-NOV-2000; 2000US-0249297.
PR   17-NOV-2000; 2000US-0249299.
PR   17-NOV-2000; 2000US-0249300.
PR   01-DEC-2000; 2000US-0250391.
PR   01-DEC-2000; 2000US-0251160.
PR   05-DEC-2000; 2000US-0251030.
PR   05-DEC-2000; 2000US-0251988.
PR   05-DEC-2000; 2000US-0256719.
PR   06-DEC-2000; 2000US-0251479.
PR   08-DEC-2000; 2000US-0251856.
PR   08-DEC-2000; 2000US-0251868.
PR   08-DEC-2000; 2000US-0251869.
PR   08-DEC-2000; 2000US-0251989.
PR   08-DEC-2000; 2000US-0251990.
PR   11-DEC-2000; 2000US-0254097.
PR   05-JAN-2001; 2001US-0259678.
XX
PA   (HUMA-) HUMAN GENOME SCI INC.
XX
PI   Rosen CA, Barash SC, Ruben SM;
XX
DR   WPI; 2001-541565/60.
XX
PS   Disclosure; SEQ ID NO 7659; 170lpp + Sequence Listing; English.
XX
PT   Nucleic acids encoding 3224 human nervous system antigen polypeptides,
PT   useful for preventing, diagnosing and/or treating nervous system
PT   cancers and metastases -
XX
CC   The invention relates to novel genes (ABA11004-ABA21534) and proteins
CC   (ABB14678-ABB18001) useful for preventing, treating or ameliorating
CC   medical conditions e.g. by protein or gene therapy. The genes are
CC   isolated from a range of human tissues disclosed in the specification.
CC   The nucleic acids, proteins, antibodies and (ant)agonists are useful
CC   in the diagnosis, treatment and prevention of: (a) cancer, e.g. breast
CC   and ovarian cancer and other cancers of the adrenal gland, bone, bone
CC   marrow, breast, gastrointestinal tract, liver, lung, or urogenital;
CC   (b) immune disorders e.g. Addison's disease, allergies, autoimmune
CC   haemolytic anaemia, autoimmune thyroiditis, diabetes mellitus, Crohn's
CC   disease, multiple sclerosis, rheumatoid arthritis and ulcerative
CC   colitis; (c) cardiovascular disorders such as myocardial ischaemias;
CC   (d) wound healing; (e) neurological diseases e.g. cerebral anoxia and
CC   epilepsy; and (f) infectious diseases such as viral, bacterial, fungal
CC   and parasitic infections.
CC   Note: The sequence data for this patent did not form part of the
CC   printed specification, but was obtained in electronic format directly
CC   from WIPO at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ   Sequence 32192 BP; 7828 A; 8820 C; 8782 G; 6762 T; 0 other;

Query Match   20.2%; Score 202; DB 22; Length 32192;
Best Local Similarity   61.6%; Pred. No. 1.9e-40;
Matches 382; Conservative   0; Mismatches 220; Indels 18; Gaps 3;

Qy   1 CCTGCCACCATGCCTGGCTAATTTTTCTTATTTTTTAGTAGAGACGAGGTTTCCCATGTTG   60
           ||||| ||||| |||||| ||| |||||||||||||||||||||||||||||||||||||
Db   6023 CCCGGCACCGCGCCCCAGCTAATTTGTGTATTTTAGTAGAGATGGGGTTTCGCATATTG   6082

Qy   61 ACCAGGCTGGTCTCGAACTCTTGACCTCAGGTGATCCGCCTGCCTCAGCCTCCCAAAGTG   120
           ||||||||||||||||||||||||||||| |||||||||||||||||||||||||||||||
Db   6083 GCCAGGCTGGTCTCAAACTCCTGACCTCAAGTGATCCACCCTGCCTCTGCCTCCCAAAGTG   6142

Qy   121 CTGGGATTATAGGCGTGAGCCCGCCGCACCCAGCCAACACATTTTTAAATACTGAAAAGTAG   180
           |||| ||| |||| |||||| |||| |||||||||| || ||| | |||| | |
Db   6143 CTGGAATTACAGGCGCATGAGCCACCGTGCCCAGCCGGGACTTTATTGATTAAAAACTTT   6202

Db   6203 TGTG------------CTGTAAATGAAAACTATCAATTAGACTCAGAGAGACACATATGTATACA   6253

Qy   181 AGGGAATAGTATATAGTGTACCCCATTTTACCCATCATCACTCAGTTTCAACAGCTGGTGACATA   240
           || ||||||| || ||||| | || | | | |||||||| | | | || | || |
Db   6254 CAATCTTACAACTCAATAATTCAATAATTTTAAAAAAACCCTACTGGGATTATTTTTAAGGAAAACC   6313

Qy   241 TTTATTTCTTCTATACCGTATACGGTACTCTCCCACTGGGATTATTTATTTTTAAGGCAAAAACC   300
           || ||| ||| | | || || ||| | ||||||||||||||| ||||||||||||||||||
Db   6314 TCACACCTGTAATCCAGCACTTTGGGAGGCCGGAGGGCGGGATCACGAGGTCAGGAGA   6373

Qy   301 AGATGACATTTTATCCCTAAATACTTTAGATAAAGGTGTTCTTTGAAAAAATCATACC   360
           ||| || | |||| |||||||||||||| | |||| ||| | ||||| | |||||
Db   6374 TCAGAGACGATCCTGGCTAACACAGTGAAACCCCGTCTCTACTAAAAATACAAAAAATTAG   6433

Qy   361 TCAGGACCAGCCTGGGCCAACATGGTGAAACCCTGTCTCTGTACTAAAAATAC-AAAAATTAG   419
           |||||| ||||||| ||| | ||||| | | |||||||| |||||||||||| |||||||
Db   6434 CCGGGCGGCAGTGGCAGGTGCCTGAAGTCCCAGCTACTCGGGAGGCTGAGGCAGGAGAATG   6493

Qy   420 CTTGGCATGGTCGTGGGCACCTGTAATCCCAGCTACTCAGGAGGCTGAGGCAGGAGAATC   479
           || ||| ||| ||| | | ||||||||||||||||| |||||||||||||||||||
Db   6494 GCGTGAACCCGGGAGGCGGAGCTTGCAGTGAGCAGAGATCGTGCCACTACACTCCAGCC   6553

Qy   480 ACTTGAATCCGGGAGGCAGAGATTGCAGTGAGCTGAGAT------TGCAGTCGAGGCT   531
           || |||||||||||| || |||||||||||| |||||| |||| | |||||||
Db   6554 GGGCGACAGACAAGACTCCATCTCAAAAACAAACAAACAAAAAAACCCT   6613

Qy   532 GGGCGACAGAGACAGAGAAATGAAAACTCTGTCTCAAAAACAAACAAACAAAAAAACCACTAT   591
           |||||||||||||| ||||||||||||||||||||||||||||||||||||||||||||||
Db   6614 ACTAAAAAATATATAAAGGAT   6633

Qy   592 ACATAAAAAATGAACAAATGAT   611
           || |||| |||| |||||||
Db   6554 GGGCGACAGACAAGACTCCATCTCAAAAACAAACAAACAAAAAAACCCT   6613

```
PR   26-JUL-2000; 2000US-0220954.
PR   14-AUG-2000; 2000US-0245518.
PR   14-AUG-2000; 2000US-0224519.
PR   14-AUG-2000; 2000US-0225213.
PR   14-AUG-2000; 2000US-0225514.
PR   14-AUG-2000; 2000US-0225266.
PR   14-AUG-2000; 2000US-0225267.
PR   14-AUG-2000; 2000US-0225268.
PR   14-AUG-2000; 2000US-0225270.
PR   14-AUG-2000; 2000US-0225447.
PR   14-AUG-2000; 2000US-0225757.
PR   14-AUG-2000; 2000US-0225758.
PR   14-AUG-2000; 2000US-0225759.
PR   18-AUG-2000; 2000US-0226279.
PR   22-AUG-2000; 2000US-0226681.
PR   22-AUG-2000; 2000US-0268868.
PR   22-AUG-2000; 2000US-0271182.
PR   23-AUG-2000; 2000US-0227009.
PR   30-AUG-2000; 2000US-0228924.
PR   01-SEP-2000; 2000US-0229987.
PR   01-SEP-2000; 2000US-0229343.
PR   01-SEP-2000; 2000US-0229344.
PR   01-SEP-2000; 2000US-0229345.
PR   05-SEP-2000; 2000US-0229509.
PR   05-SEP-2000; 2000US-0229513.
PR   06-SEP-2000; 2000US-0230437.
PR   06-SEP-2000; 2000US-0230438.
PR   08-SEP-2000; 2000US-0231242.
PR   08-SEP-2000; 2000US-0231243.
PR   08-SEP-2000; 2000US-0231244.
PR   08-SEP-2000; 2000US-0231413.
PR   08-SEP-2000; 2000US-0231414.
PR   08-SEP-2000; 2000US-0232080.
PR   08-SEP-2000; 2000US-0232081.
PR   12-SEP-2000; 2000US-0231968.
PR   14-SEP-2000; 2000US-0232397.
PR   14-SEP-2000; 2000US-0232398.
PR   14-SEP-2000; 2000US-0232399.
PR   14-SEP-2000; 2000US-0232400.
PR   14-SEP-2000; 2000US-0232401.
PR   14-SEP-2000; 2000US-0233063.
PR   14-SEP-2000; 2000US-0233064.
PR   14-SEP-2000; 2000US-0233065.
PR   21-SEP-2000; 2000US-0234223.
PR   21-SEP-2000; 2000US-0234274.
PR   25-SEP-2000; 2000US-0234997.
PR   25-SEP-2000; 2000US-0234998.
PR   26-SEP-2000; 2000US-0235484.
PR   27-SEP-2000; 2000US-0235834.
PR   27-SEP-2000; 2000US-0235836.
PR   29-SEP-2000; 2000US-0236327.
PR   29-SEP-2000; 2000US-0236367.
PR   29-SEP-2000; 2000US-0236368.
PR   29-SEP-2000; 2000US-0236369.
PR   29-SEP-2000; 2000US-0236370.
PR   02-OCT-2000; 2000US-0236802.
PR   02-OCT-2000; 2000US-0237037.
PR   02-OCT-2000; 2000US-0237038.
PR   02-OCT-2000; 2000US-0237039.
PR   02-OCT-2000; 2000US-0237040.
PR   13-OCT-2000; 2000US-0239935.
PR   13-OCT-2000; 2000US-0239937.
PR   20-OCT-2000; 2000US-0240960.
PR   20-OCT-2000; 2000US-0241221.
PR   20-OCT-2000; 2000US-0241785.
PR   20-OCT-2000; 2000US-0241786.
PR   20-OCT-2000; 2000US-0241787.
PR   20-OCT-2000; 2000US-0241808.
PR   20-OCT-2000; 2000US-0241809.
PR   20-OCT-2000; 2000US-0241826.
PR   01-NOV-2000; 2000US-0244617.
PR   08-NOV-2000; 2000US-0246474.
PR   08-NOV-2000; 2000US-0246475.
```

```
PR   08-NOV-2000; 2000US-0246476.
PR   08-NOV-2000; 2000US-0246477.
PR   08-NOV-2000; 2000US-0246478.
PR   08-NOV-2000; 2000US-0246523.
PR   08-NOV-2000; 2000US-0246524.
PR   08-NOV-2000; 2000US-0246525.
PR   08-NOV-2000; 2000US-0246526.
PR   08-NOV-2000; 2000US-0246527.
PR   08-NOV-2000; 2000US-0246528.
PR   08-NOV-2000; 2000US-0246532.
PR   08-NOV-2000; 2000US-0246609.
PR   08-NOV-2000; 2000US-0246610.
PR   08-NOV-2000; 2000US-0246611.
PR   08-NOV-2000; 2000US-0246613.
PR   17-NOV-2000; 2000US-0249207.
PR   17-NOV-2000; 2000US-0249208.
PR   17-NOV-2000; 2000US-0249209.
PR   17-NOV-2000; 2000US-0249210.
PR   17-NOV-2000; 2000US-0249211.
PR   17-NOV-2000; 2000US-0249212.
PR   17-NOV-2000; 2000US-0249213.
PR   17-NOV-2000; 2000US-0249214.
PR   17-NOV-2000; 2000US-0249215.
PR   17-NOV-2000; 2000US-0249216.
PR   17-NOV-2000; 2000US-0249217.
PR   17-NOV-2000; 2000US-0249218.
PR   17-NOV-2000; 2000US-0249244.
PR   17-NOV-2000; 2000US-0249245.
PR   17-NOV-2000; 2000US-0249264.
PR   17-NOV-2000; 2000US-0249265.
PR   17-NOV-2000; 2000US-0249297.
PR   17-NOV-2000; 2000US-0249299.
PR   17-NOV-2000; 2000US-0249300.
PR   01-DEC-2000; 2000US-0250160.
PR   01-DEC-2000; 2000US-0250391.
PR   05-DEC-2000; 2000US-0251030.
PR   05-DEC-2000; 2000US-0251479.
PR   05-DEC-2000; 2000US-0251988.
PR   06-DEC-2000; 2000US-0256719.
PR   06-DEC-2000; 2000US-0251479.
PR   08-DEC-2000; 2000US-0251856.
PR   08-DEC-2000; 2000US-0251868.
PR   08-DEC-2000; 2000US-0251869.
PR   08-DEC-2000; 2000US-0251989.
PR   08-DEC-2000; 2000US-0251990.
PR   11-DEC-2000; 2000US-0254097.
PR   05-JAN-2001; 2001US-0259678.
XX
PA   (HUMA-) HUMAN GENOME SCI INC.
XX
PI   Rosen CA, Barash SC, Ruben SM;
XX
DR   WPI; 2001-465570/50.
XX
PT   Isolated nucleic acid molecule encoding a reproductive system antigen
PT   is used in preventing, treating or ameliorating a medical condition -
XX
PS   Disclosure; SEQ ID NO 7945; 1297pp + Sequence Listing; English.
XX
CC   The present invention provides the protein and coding sequences of a
CC   number of human reproductive system related antigens. These can be used
CC   in the prevention and treatment of reproductive system disorders,
CC   including cancer. The present sequence is a genomic sequence encoding a
CC   protein of the invention.
XX
SQ   Sequence 32192 BP; 7828 A; 8820 C; 8782 G; 6762 T; 0 other;

Query Match             20.2%;  Score 202;  DB 22;  Length 32192;
Best Local Similarity   61.6%;
Matches   382;  Conservative    0;  Mismatches  220;  Indels   18;  Gaps  3;

Qy          1 CCTGCCACCATGCCTGGCCTAATTTCTTAATTTTTAGTRAGAGACGAGGTTTTGCCATGTTG 60
              ||||||| |||| || |||||||| ||||||||| ||||||||||||||| | ||||||||||| |||
Db       6023 CCCGCCACCGGCGCCCACGCGCCCAGCTAATTTTTGTATTTTTTAGTRAGAGAATGGGGTTTTGCCATATTG 6082
```

```
Qy    61  ACCAGGCTGGTCTCGAACTCTTTGACCTCAGGTGATCCGCCTGCCTCAGCCTCCCAAAGTG  120
          ||||||||||||||||||||| ||||||||||| |||||||||||| ||| |||||||||
Db  6083  GCCAGGCTGGTCTCAAACTCCTGACCTCAGGTGATCCACCTGCCTCTGCCTCCCAAAGTG  6142

Qy   121  CTGGGATTATAGGCGTGAGCCGCCGCACCCGGCCAACACATTTTTTAAATACTGAAAAGTAG  180
          ||||||||| ||||| |||||| || |||||||
Db  6143  CTGGAATTACAGGCATGAGCCACCGTGCCCAGCCCGGGACTTTTATTTGAAATTAAAAACTTT  6202

Qy   181  AGGGAAATAGTTATAGTGTACCCCATTTACCCATCACTCAGTTTCAACAGCTGGGTGACATA  240
          |||||||||||||| ||   |||   |||    ||| |
Db  6203  TGTTG--------CTGTAAATGAAAACTATCAAATTAGCTGAGAAGACAACATATGTATACA  6253

Qy   301  AGATGACATTTTATCCCTAAATACTTTAGATAAAGGTGTTCTTTGAAAAAAATCATAACC  360
          | ||| |||    |||      ||| ||
Db  6314  TCACACCTGTAATCCCAGCACTTTGGGAGGCCGAGGCGGGGATCACGAGGTCAGGAGA  6373

Qy   241  TTTATTTCTTCTATACCAGTACCGTACTCTCCCCACTGGGATTATTTTAAGGCAAAACCC  300
          | ||| |||   |||||  | |||     ||| |
Db  6254  CAATCTTACAACTCAATAATTTCAATAATTTAAAAAACCTACTAAGCCGGGCGCAGTGGC  6313

Qy   361  TCAGGACCAGCCTGGCCAACAATGGTGAAACCCTGTCTPACTAAAAATAC-AAAAATTAG  419
          | ||| ||| |||||| ||| |  ||||||||| |||||||||| - ||||||||||
Db  6374  TCGAGACGAATCCTGGCTAACACAGTGAAAACCCCGTCTCTACTAAAAATTAG  6433

Qy   420  CTTGGCATGGTGCGTGGGCACCTGTAATCCCAGCTACTCAGGAAGGCTGAGGCAGGAGAATC  479
          | || ||||||||||||| |||||||||||||||||||||||||||||||||||||||
Db  6434  CCGGGCGCAGTGGCGTGGGCTGCCTGAAGTCCCAGCTACTCGGGAAGCTGAGGCAGGAGAATG  6493

Qy   480  ACTTGAATCCGGGAAGCAGAGAGATTGCAGTGAGCTGAGAT--------TGCAGTCGAGCCT  531
          | ||| |||||||||||||||||||| ||||||||||||||||        || ||||||
Db  6494  GCGTGAACCCGGGAGGCGGAGCTTGCAGTGAGCAGATCGTTGCCACTACACTCCAGCCC  6553

Qy   532  GGGCGACAGAGCAGAGAAATGAAACTCTGTCTCAAAAAAACAAACAAAAAAACCACTAT  591
          |||||||||||| ||||  |||  ||||| |||||||||||||||||||| ||||
Db  6554  GGGCGACAGAGCAAGAGACTCCATCTCAAAAACAAACAAACAAACAAAAAAAACCCT  6613

Qy   592  ACATAAAAATGACAAATGAT  611
          || |||||  || |||
Db  6614  ACTAAAAATATATAAAGGAT  6633
```

GenCore version 5.1.4_p5_4578
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 29, 2003, 16:20:14 ; Search time 1834.16 seconds
                                           (without alignments)
                                           8829.908 Million cell updates/sec

Title: US-09-691-219-3_COPY_1_1000
Perfect score: 1000
Sequence: 1 cctgccaccatgcctggcta.........caagaggcctgcaccctga 1000

Scoring table: IDENTITY_NUC
               Gapop 10.0 , Gapext 1.0

Searched: 16154066 seqs, 809774376 residues

Total number of hits satisfying chosen parameters: 32308132

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 08
                 Maximum Match 100%
                 Listing first 45 summaries

Database : EST:*
    1: em_estba:*
    2: em_esthum:*
    3: em_estin:*
    4: em_estmu:*
    5: em_estov:*
    6: em_estpl:*
    7: em_estro:*
    8: em_htc:*
    9: em_est1:*
   10: gb_est2:*
   11: gb_htc:*
   12: gb_est3:*
   13: gb_est4:*
   14: gb_est5:*
   15: em_estom:*
   16: em_estfun:*
   17: gb_gss:*
   18: em_gss_hum:*
   19: em_gss_inv:*
   20: em_gss_pln:*
   21: em_gss_vrt:*
   22: em_gss_fun:*
   23: em_gss_mam:*
   24: em_gss_mus:*
   25: em_gss_other:*
   26: em_gss_pro:*
   27: em_gss_rod:*

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
	1	199.2	19.9	967	12	BE897079	BE897079 601439635
	2	171	17.1	701	12	BE744242	BE744242 601576428
	3	162	16.2	652	12	BG116323	BG116323 602318619
c	4	160.6	16.1	672	17	AG102362	AG102362 Pan trogl
c	5	160.6	16.1	696	17	AG106956	AG106956 Pan trogl
	6	160.2	16.0	684	17	AG088170	AG088170 Pan trogl
c	7	159.8	16.0	523	17	A2517953	A2517953 RPCI-11-3
	8	158.4	15.8	449	17	A2773470	A2773470 UP_524-15
	9	157	15.7	394	14	BM686844	BM686844 UI-E-CR1-
	10	156.8	15.7	390	9	AA229935	AA229935 nc51g10.r
c	11	156.6	15.7	664	13	BI823844	BI823844 603039006
c	12	156.2	15.6	502	17	AQ588276	AQ588276 CITBI-E1-
c	13	156.2	15.6	566	17	AQ549290	AQ549290 RPCI-11-4
c	14	156.2	15.6	682	14	BM970657	BM970657 UI-CF-EC1
c	15	156	15.6	702	12	BF678427	BF678427 602085862
c	16	155.8	15.6	327	9	AA551519	AA551519 n108f10.s
c	17	155.6	15.6	448	17	B15808	B15808 346C24.TP C
c	18	155.6	15.6	992	14	BM914542	BM914542 AGENCOURT
c	19	155.6	15.6	585	17	AQ549426	AQ549426 RPCI-11-4
c	20	154.4	15.4	564	17	AQ352599	AQ352599 CITBI-E1-
c	21	154	15.4	458	12	BF826787	BF826787 MR2-HN003
c	22	154	15.4	587	17	AQ275559	AQ275559 RPCI-5-88
c	23	154	15.4	678	17	AQ388600	AQ388600 RPCI11-15
.	24	153.8	15.4	591	9	AL079812	AL079812 DKFZp434P
c	25	153.8	15.4	1638	11	BC015230	BC015230 Homo sapi
c	26	153.6	15.4	447	15	AQ045665	AQ045665 CIT-HSP-2
c	27	153.4	15.3	478	12	BF832503	BF832503 PM3-HT092
c	28	153.4	15.3	699	17	B94348	B94348 CIT-HSP-217
c	29	153.4	15.3	1103	12	BG116267	BG116267 60231851
c	30	153.2	15.3	525	17	B56395	B56395 CIT-HSP-200
c	31	153.2	15.3	750	14	BM678642	BM678642 UI-E-EC0-
c	32	152.6	15.3	718	17	AG011207	AG011207 Homo sapl
c	33	152	15.2	344	14	BM686853	BM686853 UI-E-CR1-
c	34	152	15.2	721	12	BG536900	BG536900 60256664
c	35	151.6	15.2	317	17	A2521678	A2521678 RPCI-11-1
c	36	151.4	15.1	503	14	BQ181246	BQ181246 UI-H-EU0-
c	37	151.2	15.1	548	17	AQ609133	AQ609133 HS_5373_B
c	38	151.2	15.1	420	7	B41620	B41620 HS-1054-A1-
c	39	151.2	15.1	482	9	AA565733	AA565733 nj23g05.s
c	40	151.2	15.1	536	17	AQ415831	AQ415831 RPCI-11-2
c	41	151.2	15.1	549	17	AQ318141	AQ318141 RPCI11-94
c	42	150.8	15.1	397	10	AW020150	AW020150 df05a09.y
c	43	150.8	15.1	537	10	AQ322719	AQ322719 RPCI11-10
c	44	150.8	15.1	3777	17	AF101957	AF101957 RPCI11-10
c	45	150.4	15.0	466	10	AW955724	AW955724 EST367794

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

ALIGNMENTS

RESULT 1
BE897079
LOCUS   BE897079   967 bp   mRNA   linear   EST 20-OCT-2000
DEFINITION 601439635F1 NIH_MGC_72 Homo sapiens cDNA clone IMAGE:3924489 5',
            mRNA sequence.
ACCESSION   BE897079
VERSION   BE897079.1 GI:10362159
KEYWORDS   EST.
SOURCE   human.
ORGANISM   Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1 (bases 1 to 967)
AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
TITLE   National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL   Unpublished (1999)
COMMENT   Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: ATCC/DCTD/DTP
            cDNA Library Preparation: Life Technologies, Inc.
            cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
            DNA Sequencing by: Incyte Genomics, Inc.
            Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLAM9762 row: f column: 10
            High quality sequence stop: 584.
FEATURES   Location/Qualifiers
source   1..967

```
                    /organism="Homo sapiens"
                    /db_xref="taxon:9606"
                    /clone="IMAGE:3924489"
                    /clone_lib="NIH_MGC_72"
                    /tissue_type="melanotic melanoma"
                    /lab_host="DH10B (phage-resistant)"
                    /note="Organ: skin; Vector: pCMV-SPORT6; Site_1: NotI;
                    Site_2: SalI; Cloned unidirectionally. Primer: oligo dT.
                    Average insert size 2 kb. Library constructed by Life
                    Technologies."
BASE COUNT    240 a   240 c   251 g   236 t
ORIGIN

Query Match   19.9%;   Score 199.2;   DB 12;   Length 967;
Best Local Similarity   64.7%;   Pred. No. 1.8e-22;
Matches 370;   Conservative  0;   Mismatches 178;   Indels 24;   Gaps 4;

Qy   11 TGCCTGGCTAATTTTTGTTATTTTGTTTTAGTAGAGACGAGGGTTTTGCCATGTTGACCAGGCTGG   70
         ||||  || ||| ||||| ||||||||||||||||||||||| ||| ||||||||| |||||||||
Db    3 TGCCCTGTAATATTTGTTATTTTGTTTTGGTAGAGACGAGGTTTTTCGCCATGTTGGCCAGGCTGG   62

Qy   71 TCTCGAACTCTTGACCTCAGGTGATCCGCCCTGCCTCAGCCTCCCAAAGTGCTGGGATTAT  130
         ||| |||||| ||||||||||||||||  ||||||||| |||||||||||||||||||||
Db   63 TCTTGAACTCCTGACCTCAGGTGATCCAACCTGCCTCGGCCTCCCAAAGTGCTGGGATTAA  122

Qy  131 AGGCGTGAGCCGCCGCGCCCAGCCAACATTTTTTAAAATACTGAAAAGTAGAGGGAATAGT  190
         |||| |||||| ||| |||  |||    ||||| | |||  |||||||| ||
Db  123 AGGCATGAGCCACCATGCCTGGCCTCATTGTAGTTTCGATT---------TGCATT      169

Qy  191 TATAGTGTACCCCATTTTACCCATCACTCAGTTTCAACAGCTGGTGACATATTTATTTCTT  250
         | |||| | ||    || ||  |                     |     |||    |||
Db  170 TCTCTGATGATCAGTGAAGTTAAGCATCTTTTTATATCCGTTGGAGATTTAAATGTCTT    229

Qy  251 CTATACCAGTACCGTACTCTCCCCACTGGGATTATTTGAAGGCAAAAACCCAGATGACATT  310
         | ||  |  ||| |  |                    ||  |||    | |   ||||
Db  230 ATTTTATA------AAATGTCTGTTCTCGGGCCTCGGGGTGTTGGTGTGTGGGTGGTTTGGGC  283

Qy  311 TTTATCCCTTAAATACTTTAGATAAAGGTGTTCTTTGAAAAAATCATAACCTCAGGACCAG  370
              |       ||  |  ||           | |                   |||||||
Db  284 AGCACTTGGGAGAGGCTGAAGTGGGTGTGATCATCTG----AGGTCAGTGCTTGCAGGACCAG  339

Qy  371 CCTGGGCCAACAATGGTGAAAACCCTGTCTGTCTACTAAAAATAACAAAATTAGCTGGGCATGGT  430
         ||||| ||||| ||||||||| ||||  |||  |  ||||||| ||||| |||||||||||||
Db  340 CCTGGACAATATGGTGAAACCCCAATCTCTGCTAAAAATACAAAAAATTGGCTGGGCATGGT     399

Qy  431 CGTGGGGCACCTGTAAATCCCAGCTACTCAGGGAAGCTGAGGCAGGAGAATCACTTGAATCCG  490
         |||||| | ||||||||||||||||||||||| |||||||||| |||||||| |||||| |
Db  400 GGTGGGCGGGCCCTGTAAATCCCAGCTACTCAGGGAGGCTGAGGCATGAGAATTGCTTGAACCCA  459

Qy  491 GGAAGCAGAGATTGCAGTGAGCTGAGATTGCAGTGAGCCGAGACGAGACAGA-AA  549
         ||| || ||| |||||||||| |||||| |||  ||  |||| |  |
Db  460 GGAGGCGGAGGTTGCAGTGAGCCGAGATCGCACTCCAGCCCGGGTGATAGAT      519

Qy  550 TGAAAACTCTGTCTCTCAAAAACAAAAACAAAACAAAA  581
         |||  ||||||||| |||||||||||||||||||||||
Db  520 TGAGACTCTCTCTCAGAAAAACAAAAACAAAACAAAA  551
```

```
RESULT 2
LOCUS       BE744242     701 bp    mRNA     linear    EST 15-SEP-2000
DEFINITION  601576428F1 NIH_MGC_9 Homo sapiens cDNA clone IMAGE:3837367 5',
            mRNA sequence.
ACCESSION   BE744242
VERSION     BE744242.1  GI:10158234
KEYWORDS    EST.
SOURCE      human.
ORGANISM    Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 701)
AUTHORS     NIH-MGC http://mgc.nci.nih.gov/,
TITLE       National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL     Unpublished (1999)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: DCTD/DFP
            cDNA Library Preparation: Ling Hong/Rubin Laboratory
            cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
            DNA Sequencing by: Incyte Genomics, Inc.
            Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at: Image.llnl.gov
            Plate: LLCM522 row: h column: 08
            High quality sequence stop: 666.

FEATURES             Location/Qualifiers
     source          1..701
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:3837367"
                     /clone_lib="NIH_MGC_9"
                     /tissue_type="adenocarcinoma cell line"
                     /lab_host="DH10B (phage-resistant)"
                     /note="Organ: ovary; Vector: pOTB7; Site_1: XhoI; Site_2:
                     EcoRI; cDNA made by oligo-dT priming. Directionally
                     cloned into EcoRI/XhoI sites using the following 5'
                     adaptor: GGCACGAG(G). Size-selected >500bp for average
                     insert size 1.8kb. Library constructed by Ling Hong in
                     the laboratory of Gerald M. Rubin (University of
                     California, Berkeley) using ZAP-cDNA synthesis kit
                     (Stratagene) and Superscript II RT (Life Technologies)."
BASE COUNT    195 a   176 c   170 g   160 t
ORIGIN

Query Match   17.1%;   Score 171;   DB 12;   Length 701;
Best Local Similarity   61.8%;   Pred. No. 6e-18;
Matches 370;   Conservative  0;   Mismatches 200;   Indels 29;   Gaps 5;

Qy    4 GCCACCATGCCTGGCTAATTTTCTTATTTTTAGTAGAGACGAGGGTTTTGCCATGTTGACC   63
         ||||| |  || |||| |||||||  | ||||||||||||||   ||| |||| |||||||
Db  113 GCCACCTTGCCCCGGC-AAATTTTTTGTGTATTTTTAGTAGAGACGCTGTGTTCACCGTGTTGGGC  171

Qy   64 AGGCTGGTCTCGAACTCTTGACCTCAGGTGATCCGCCCGCTCCGCCTCCCAAAGTGCTG  123
         |||||||||||  |||| |||  ||| |||||||  ||    | ||| ||||||||||||
Db  172 AGGCTGGTCTCCAACTCCTGAACCTCATGTGATCCACCCACCTCAGCCTTCCCAAAGTGCTG  231

Qy  124 GGATTATAGGCGTGAGCCGCCGCACCCAGCCAACATTTTTTAAATACTGAAAAGTAGAGG  183
         |||||| |||||||||||||||||||| |||||||  ||   |  |||
Db  232 GGATTACAGGCGTGAGCCACCGCACCCGCCAACACTTGCTTTACCTTTTATATTATTT-----  286

Qy  184 GAATAGTTATAGTGTACCCCATTTACCCATCACTCAGTTTCAACAGCTGGTGACATATTT  243
           |  |    |      ||| ||   |             |  | |||     |    |||
Db  287 ------TTGAAGTAAAATCCAAGATAATCATTTCATCTGTAAACATTAACCATACACAGTATT  340

Qy  244 ATTTCTTCTATACCAGTACCGTACTCTCCCCACTGGGATTATTTTAAGGCAAAACCCAGA  303
         || |  | |||  |  ||                     |||  | |||    |||||
Db  341 ATTATACATGCTTTAAAAAGTAGTAAAACCCAGGCTG----------ACGAGGTGACTCA  389

Qy  304 TGACATTTTATCCCTAAATACTTTAGATAAAGGTGTTCTTTGAAAAAATCATAACCTCA  363
         ||||  | | |||| |  |                                   |  |
Db  390 CGCCTGTAATCCCAGCACTTTGGGAGGCCAAGGCGGGAACACGAGGTCAGGAGATCG    449

Qy  364 GGACCAGCCTGGCCAACATGGTGAAACCCTGTCTCTACTAAAAATACA-------AAAATTA  418
         |||||| |||||| ||||||||||| ||| || |||||||||||||||       |||||||
Db  450 AGACCATCCTGGCTAACATGGTGAAAACCCCGTTTCTACTAAAAATACAAAAAGAAAAATTA   509

Qy  419 GCTTGGGCATGGTGGCTGTGGCGCACCTGTAAATCCCAGCTACTCAGGAGGCTGAGGCAGGAGAAT  478
         ||  ||||||||||||| | |||  |  ||||| ||||||| || ||||||||||||||||||||
Db  510 GCCGGGCATGGTGGCGGGATCGCGTGGTAGTCCCAGCTGCTCGGGAGGCTGAGGCAGGAGAAT     569

Qy  479 CACTTGAATCCGGGAAGCAGAGATTGCAGTGAGCTGAGATTGCAGTGAGCTCGGGCGAC  538
           | ||||  ||||||| | ||| |||  ||||||||||| |    ||   |||||| ||
Db  570 GGCATGAACCCGGGAGGCGGAGGTTGCTTGCAGTGAGCTGAGATCATGCCACTGCACTCCAGCC  629

Qy  539 AGAGACAGA-AATGAAACTTGTGTCTCAAAAACAAAAACAAAAACAACAAACACTATACATA  596
         ||||||||| | |||  | |  | |||||||| |   |||    ||| |    |  | |||
Db  630 TGGGAGACAGAGCGAGACCGACTCTGTCTCACAAAAAAAAAAAAAAAAAAAAACAACAAAAAAAA  688
```

```
RESULT 3
LOCUS       BG116323          652 bp    mRNA    linear   EST 30-JAN-2001
DEFINITION  602318619F1 NIH_MGC_88 Homo sapiens cDNA clone IMAGE:4418692 5',
            mRNA sequence.
ACCESSION   BG116323
VERSION     BG116323.1  GI:12609829
KEYWORDS    EST.
SOURCE      human.
ORGANISM    Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 652)
AUTHORS     NIH-MGC http://mgc.nci.nih.gov/.
TITLE       National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL     Unpublished (1999)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: ATCC
            cDNA Library Preparation: Life Technologies, Inc.
            cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
            DNA Sequencing by: Incyte Genomics, Inc.
            Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLAM10153 row: f column: 05
            High quality sequence stop: 630.
FEATURES             Location/Qualifiers
     source          1..652
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:4418692"
                     /clone_lib="NIH_MGC_88"
                     /tissue_type="duodenal adenocarcinoma, cell line"
                     /note="Organ: small intestine; Vector: pCMV-SPORT6;
                     Site_1: NotI; Site_2: SalI; Cloned unidirectionally.
                     oligo-dT primed. Average insert size 1.767 kb. Library
                     enriched for full-length clones and constructed by Life
                     Technologies. Note: this is a NIH_MGC Library."
                     /lab_host="DH10B (phage-resistant)"
BASE COUNT      163 a    170 c    175 g    144 t
ORIGIN

Query Match             16.2%;  Score 162;  DB 12;  Length 652;
  Best Local Similarity   58.7%;
  Matches  345;  Conservative    0;  Mismatches  225;  Indels   18;  Gaps    3;

Qy    1 CCTGCCACCATGCCTGGGCTAATTTTCTTATTTTTAGTAGAGACGGAGGTTTTGCCATGTTG 60
          || ||||||||||||| |||||||||| ||| || ||||||  |||||||||||
Db   75 CCACCACCGCACCTGGCTAATTTTTAAATTTTTATARAGAGATGGAGTCTCACTGTGTTG 134

Qy   61 ACCAGGCTGGTCTCGGACTCTTGACCTCAGGTGATCGCGCCTGCCTCAGCC-TCCCAAAGT 119
          |||||||||| | |||||| | |||||||| |||||  |||||  |||| || |||
Db  135 CCCAGGCTGGTGTTGAACTCTCGGCCTCAAGTGATCTCCCAGCCTTCCTGAAGT 194

Qy  120 GCTGGGATTATAGGCGTGAGCCCGCCGCGCCCAGCCACATTTTTTAAATAC---TGAAAA 176
          |||||||||| ||||| |||| ||  || |||| |  | ||||  || |  ||||||
Db  195 GCTGGGATTACAGGCATGAGTCGCTGCCACCCAGTCTGGTTCTTAACAAGCCTATGCAAA 254

Qy  177 GTAGAGGGAATAGTTATAGTGTGTACCCCATTACCCATCACTCAGTTGGGTGA 236
          | || | | ||  |||| |||| |  | |||||| ||| || ||||||||||
Db  255 ATTTAAAATCTCTGTTGCCATCGTCAGTCTCGAAATCTCAAAAAAAATTACCTTTTGACCTGT 314

Qy  237 CATATTTATTTCTTCTATACCAGTACCGTACTCTCCCACTGGGATTATTTTAAGGCAA 296
          ||| ||| | ||||| || ||| |||  ||| || || |||||| | |||||||||
Db  315 CAGGCAAATTGGTTTAAAACATTCAAAAATGTCTTGGCTGGCACGGTGGCTCACGCCTGT 374

Qy  297 ACCCAGATGACATTTATCCCTAAATACTTTAGATAAAAGGTGTTCTTTGAAAAAAATCAT 356
          || ||||| || | ||| | ||||| || | | |||||||| | |||||||||| |||
Db  375 AATCCCAGCACTTGGGGAGGCTGAGGCGGGATGGATCACG---------AGGTCAG 420
```

```
RESULT 4
LOCUS       AG102362          672 bp    DNA     linear   GSS 03-NOV-2001
DEFINITION  Pan troglodytes DNA, clone: PTB-105K09.R, genomic survey sequence.
ACCESSION   AG102362
VERSION     AG102362.1  GI:16722879
KEYWORDS    GSS.
SOURCE      Pan troglodytes male lymphoblast DNA, clone_lib:PTB Chimpanzee Male
            BAC Library clone:PTB-105K09.R.
ORGANISM    Pan troglodytes
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Pan.
REFERENCE   1
AUTHORS     Fujiyama,A., Hattori,M., Toyoda,A., Taylor,T.D., Yada,T.,
            Totoki,Y., Watanabe,H. and Sakaki,Y.
TITLE       BAC end sequences of Library PTB
JOURNAL     Unpublished
REFERENCE   2  (bases 1 to 672)
AUTHORS     Fujiyama,A., Hattori,M., Toyoda,A., Taylor,T.D., Yada,T.,
            Totoki,Y., Watanabe,H. and Sakaki,Y.
TITLE       Direct Submission
JOURNAL     Submitted (02-AUG-2001) Asao Fujiyama, The Institute of Physical
            and Chemical Research (RIKEN), Genomic Sciences Center (GSC);
            1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan
            (E-mail:chimpbac@gsc.riken.go.jp, URL:http://hgp.gsc.riken.go.jp/,
            Tel:81-45-503-9111, Fax:81-45-503-9170)
COMMENT     Clones are derived from the chimpanzee BAC library PTB This BAC end
            was generated during the R&D process and may have higher chance of
            clone tracking errors.
FEATURES             Location/Qualifiers
     source          1..672
                     /organism="Pan troglodytes"
                     /db_xref="taxon:9598"
                     /clone="PTB-105K09.R"
                     /sex="male"
                     /cell_type="lymphoblast"
                     /clone_lib="PTB Chimpanzee Male BAC Library"
              LIBRARY
              Vector    : pKS145
              R.Site 1  : SacI
              R.Site 2  : SacI.
              PRIMERS
              Sequencing: M13Rev
BASE COUNT      209 a    145 c    155 g    163 t
ORIGIN

Query Match             16.1%;  Score 160.6;  DB 17;  Length 672;
  Best Local Similarity   72.5%;  Pred. No. 2.7e-16;
  Matches  208;  Conservative    0;  Mismatches   79;  Indels    0;  Gaps    0;

Qy  357 AACCTCAGGACCAGCCTGGCCAACATGGTGAAAACCCTGTCTGTACTAAAAATACAAAAAT 416
          || ||||| ||| ||||||| |||| ||||||| ||| |||| | || |||||| |||||
Db  421 GAGATGACACCATCCTGGCTNAACACGGTGAAACCCGTCTCGTGTGAAAATATATAAAAT 480

Qy  417 TAGCTTGGCATGGTGGCGTGGGCACCGTAATCCACGCTACTCAGGAAGCTGAGGCAGGAGA 476
          |||| |||||||||||||  ||| ||||||| ||||| | |||||||||||||||  |||||
Db  481 TAGCCGGGCATGGTGGCGGGCGCCTGTAGTCCCGGCTACTGGGAGGCTGAGGCGGGAGA 540

Qy  477 ATCACTTGAATCCCGGGAAGCAGGAGGTTGCAGTGAGCTGAGATTGCAGTCAGCCTGGGCG 536
          || | ||||| ||||||||||||| |  |||||||||| |||| | | ||  ||||||||
Db  541 ATGGCATGAACCCCGGGAGGCAGAGCTTGCAGTGAGCCGAGATCGCGCCACTGCACTCCAGC 600

Qy  537 ACAGGACACGAAAATGAAAACTCTGTCTCAAAAAACAAACAAACAAACAAAAAA 584
          || | | |  || || |||| ||||||||||||| |||| |||||| |||||
Db  601 CAGGCCAGAGAGCGAGAGTCTTGTCTCAAAAAACAAAAACAAACACAAA 648

Qy  348 AAAAATCATAAACCTCAGGAGACCAGCCTGGCCAACACAGGTGAAAACCCTGTCTGTACTAAAAA 407
          ||||||||| | ||||| ||| ||||||||||||| ||  ||||| |||| |||| | ||||||
Db   61 AGAGCTCAGGAGTTCGAGACCAGCCTGGCCAACATGGTGAAACCCCGTCTCTACTAAAAA 120

Qy  408 TACAAAAAATTAGCCTTGGCATGGTGGTCGTGGGCACCCTGTAATCCCAGCTACTCAGGAGGCTGA 467
          |||||||||||||||||||||||||||| |||||||| |||||||||||||||||||||| ||||
```

RESULT 5

```
LOCUS       AG106956                  696 bp    DNA     linear   GSS 03-NOV-2001
DEFINITION  Pan troglodytes DNA, clone: PTB-111K02.R, genomic survey sequence.
ACCESSION   AG106956
VERSION     AG106956.1  GI:16727474
KEYWORDS    GSS.
SOURCE      Pan troglodytes
  ORGANISM  Pan troglodytes
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Pan.
REFERENCE   1  (bases 1 to 696)
  AUTHORS   Fujiyama,A., Hattori,M., Toyoda,A., Taylor,T.D., Yada,T.,
            Totoki,Y., Watanabe,H. and Sakaki,Y.
  TITLE     BAC end sequences of Library PTB
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 696)
  AUTHORS   Fujiyama,A., Hattori,M., Toyoda,A., Taylor,T.D., Yada,T.,
            Totoki,Y., Watanabe,H. and Sakaki,Y.
  TITLE     Direct Submission
  JOURNAL   Submitted (02-AUG-2001) Asao Fujiyama, The Institute of Physical
            and Chemical Research (RIKEN), Genomic Sciences Center (GSC);
            1-7-22 Suehiro-chou, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan
            (E-mail:chimpbes@gsc.riken.go.jp, Fax:81-45-503-9170)
            Tel:81-45-503-9111, Fax:81-45-503-9170, URL:http://hgp.gsc.riken.go.jp/,
COMMENT     Clones are derived from the chimpanzee BAC library PTB This BAC end
            was generated from the chimpanzee BAC library PTB This BAC end
            clone tracking errors.
FEATURES             Location/Qualifiers
     source          1..696
                     /organism="Pan troglodytes"
                     /db_xref="taxon:9598"
                     /clone="PTB-111K02.R"
                     /sex="male"
                     /cell_type="lymphoblast"
                     /clone_lib="PTB Chimpanzee Male BAC Library"
     PRIMERS
                     Sequencing: M13Rev
     LIBRARY
                     Vector    : pKS145
                     R.Site 1  : SacI.
                     R.Site 2  : SacI.
BASE COUNT      157 a    172 c    165 g    199 t    3 others
ORIGIN
```

```
Query Match         16.1%;   Score 160.6;  DB 17;   Length 696;
Best Local Similarity  60.0%;   Pred. No. 2.6e-16;
Matches  346;  Conservative    0;   Mismatches  189;   Indels  42;   Gaps   3;
```

```
Qy    16 GGCTAATTTTTCTTATTTTTTTAGTAGAGACGAGGTTTGCCATGTTGACCAGGCTGGTCTCG  75
         |||| |||||||| ||| ||| |||||||||||||||||||| |||||||||||||||||
Db   695 GGTAATTTTTGTGGTTTCTTTGTTAGAGGCAGGGTTTCACCATGTTGCCCAGTGTGGTCTCCA  636

Qy    76 AACTCTTGACCTCAGGTGATCCGCGTCCGCCTCAGCCTCCCAAAGTGCTGGGATTATAGGCG  135
         |||| ||| | |||||||||| ||| ||||| |||||||||||| || |||||||| |||||
Db   635 AACTCGTGGGCTCAAGTCATGTACCTGCTTCAGCCTGCCAAAGTGCTGGGATTACAGGTG  576
```

RESULT 6

```
LOCUS       AG088170                  684 bp    DNA     linear   GSS 03-NOV-2001
DEFINITION  Pan troglodytes DNA, clone: PTB-087A17.R, genomic survey sequence.
ACCESSION   AG088170
VERSION     AG088170.1  GI:16639972
KEYWORDS    GSS.
SOURCE      Pan troglodytes
  ORGANISM  Pan troglodytes
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Pan.
REFERENCE   1  (bases 1 to 684)
  AUTHORS   Fujiyama,A., Hattori,M., Toyoda,A., Taylor,T.D., Yada,T.,
            Totoki,Y., Watanabe,H. and Sakaki,Y.
  TITLE     BAC end sequences of Library PTB
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 684)
  AUTHORS   Fujiyama,A., Hattori,M., Toyoda,A., Taylor,T.D., Yada,T.,
            Totoki,Y., Watanabe,H. and Sakaki,Y.
  TITLE     Direct Submission
  JOURNAL   Submitted (02-AUG-2001) Asao Fujiyama, The Institute of Physical
            and Chemical Research (RIKEN), Genomic Sciences Center (GSC);
            1-7-22 Suehiro-chou, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan
            (E-mail:chimpbes@gsc.riken.go.jp, Fax:81-45-503-9170)
            Tel:81-45-503-9111, Fax:81-45-503-9170, URL:http://hgp.gsc.riken.go.jp/,
COMMENT     Clones are derived from the chimpanzee BAC library PTB This BAC end
            was generated during the R&D process and may have higher chance of
            clone tracking errors.
FEATURES             Location/Qualifiers
     source          1..684
                     /organism="Pan troglodytes"
                     /db_xref="taxon:9598"
                     /clone="PTB-087A17.R"
     PRIMERS
                     Sequencing: M13Rev
     LIBRARY
                     Vector    : pKS145
                     R.Site 1  : SacI.
                     R.Site 2  : SacI.
```

```
Qy   136 TGACGCGCGCACCCAGCCAACACATTTTTTAAATACTGAAAAGTAGAGGAAATAGTTATAG  195
         |||||| ||| ||| ||||| |||||||||  |||||| ||| | ||||| ||||| |||
Db   575 TGAGCCACTGCACTCGGGCGCCATTAT--------------------AAATTGTTTTAG  539

Qy   196 TGTACCCCACATTTACCCATACTCACTCAGTTTCAACACTGTTGACATATTTTTCTTCTATA  255
         ||| | ||| ||| |||| |||  |||||||| |||| |||| ||| ||| ||||  |||
Db   538 AGCACTGCGGCGCCAAGCAAAATGTATCTATGGGGTCATGTCAAAACCACTATCAATATG  479

Qy   256 CCAGTACCGTACTCTCCCCACTGGGATTATTTTTAGGCAAAACCCAGATGACATTTTATC  315
         || ||  |||||  ||||||| ||||| ||||  |||| |||| || |||| ||||||
Db   478 AGACTCCTGGCTGGGCCCTGGGTGG----------CTCATACCTGTTAATCCTA  437

Qy   316 CCTAAATACTTTAGATAAAGGTGTTCTTTGAAAAAATCATAACCTCAGGACCAGCCTGG  375
         || |||| |||||| || || ||||||||| ||||| ||||| ||| || ||||| |||
Db   436 GCCTTATAGGAGACCAAGGTGGGGCGTGATTAACCTPAAGGTCAGGAGGCTCAAGACCAGCGTGG  377

Qy   376 CCAACACATGGTGAAAACCCTGTCTCTPGTACTAAAAAATAACAAAATTAGCTTGGCGTCGTGG  435
         ||| | |||| ||||| |||||| |||| | ||||||| ||||| |||| |||||| || |||
Db   376 CCAACACTGGCAAAAACCCGTCTTTPACTACAAAACAACAAAAACTTAGCTGGGCGTGGTGGCAC  317

Qy   436 GCACCTGTPAATCCCCAGCTYACTCAGGAAGCTGAGGCAGGAGGAGAATCACTTGAAATCCGGGAAG  495
         |||| ||||| ||||||| || ||| ||||||| ||| |||||||||||||||| |  | ||||||
Db   316 ATGCCTGTAATCCCAGCTACTTGGGAGGCTAAGACAGAAGAAAATGCTTGAAACCAGGAGG  257

Qy   496 CAGAGATTGCAGTGAGCTGAGATTGCCAGTCGAGCCTGGGCGACAGAGCAAGACTCCATCT  554
         || ||||||| ||||||||||| ||| |||| ||||| || |||||||| || || |||
Db   256 CAGAGGTTGCAGTGAGTCGAGATCGCGCCACTCCACTCCAGCCTGGGCAACAGAGCGAGA  197

Qy   555 CTCTGTCTCAAAAAACAAACAACAAAACAAAAAAAAACCACTAT  591
         |||||||| ||||||||||||||||||||||||||| ||||||
Db   196 CTCCTTCTCAAAAAAAAAAAAAAAAAAAAAAAAAAAAGAAT  160
```

```
                /sex="male"
                /cell_type="lymphoblast"
                /clone_lib="PTB Chimpanzee Male BAC Library"
BASE COUNT      209 a    158 c    162 g    155 t
ORIGIN

Query Match     16.0%;   Score 160.2;  DB 17;   Length 684;
Best Local Similarity    60.1%;   Pred. No. 3.1e-16;
Matches 325;   Conservative  0;   Mismatches 203;   Indels 13;   Gaps 3;

Qy   63 CAGGCTGGTCGTCGAACTCTTGACCTCAGGTGATCCGCCCGCCTCCCAAAGTGCT 122
        ||||| | |||  |||||||| |||||||||| |  ||||||||| |||||||||||
Db   33 CCGCGCGCTAGCAGTGCTCGAGAGGTCAACGGAGTCCGCCAGCCTCAGTCTCCTAAAGTGCT 92

Qy  123 GGGGATTATAGGCGTGAGCCGCCGCGCCACCCAACACATTTTTTAAATACTGAAAAGTAGAG 182
        |||||||| |||  | ||||| ||| |||  | ||| | ||||||     |||| |||| |
Db   93 GGGGATTACAGGTGCGAGCCCACCGTGCCCAGCCTCATATATTTT-----TATATTCATTTT 146

Qy  183 GGAATAGTTATTAGTGTACCCCATTTACCCATCACTCAGTTTCAACAGCGTGGTGACATATT 242
        | | | |  |  | | |||  | |||| ||| | | ||||| || || |  |   | |||
Db  147 GAAAAAAAACAAAACATACAAAAATATTAGATAAGCATTTAGATCACAAGTTCCCAAACTGG 206

Qy  243 TATTTCTTCTATACCAGTACCGTACTCTCCCCACTGGGATTATTTTATGTCAAAACCCAG 302
        | ||| |   |  |  | |  | | |  |||| |||||| |||||   | ||||||| |
Db  207 TCTTTG-----TAGCAGCAATTTAGGAAATCACCTGGGTAGTTTTTAAAAATACGTATTCT 261

Qy  303 ATGACATTTTATCCCTAAATACTTTAGATAAAAGGTGTTCTTTGAAAAAAATCATAACCTC 362
        | || |  | | |  || | | | || | |  ||| |  | | |||||||| || ||||||
Db  262 GCCGGGCACGGTGGCTCACTTGGGAGGCCGAGGCGGGCGGATCATGAGGTCAGGAGATC 321

Qy  363 AGGACCAGCCTCGGCCAACATAGGTGAAAACCCTGTCTGTACTAAAAATAC--AAAAATTAGC 420
        ||| |||| || ||||||||  ||||||| || | | ||||||||||| | || ||||||||||
Db  322 GAGACCATCCTGGCCAACACGGTGAAACCCCATCTCTACTAAAAATACAAAAAATTAGC 381

Qy  421 TTGGCATGGTGGTGGGCGCACCTGTAATCCCAGCTACTCAGGGAGGCTGAGGCAGGAGAATCA 480
        | |||||||||||||||||  |||||||||||||||||| |||||||||||||||||||||| |
Db  382 CGGGCCATGGTGGTGGGCGCCTGTAATCCCAGCTACTCGGGAGGCTGAGGCAGGAGAATGG 441

Qy  481 CTTGAATCCGGGAGGCAGAGGATTGCAGTGAGCTGAGATTGCAGTGCGAGCCGCTGGGCGACAG 540
        |||||||| ||||||| ||||  |||||||||| ||||| ||    || ||| | |   ||| |
Db  442 CATGAATCCAGGAGGCGGAGGCTTGCAGTGAGCCGAGATCGCGCCACTGCACTCCAGCCTG 501

Qy  541 AGACAGAAATGAAACTCTGTCTCAAAAAACAAAACAAAACCCACTATACATAAAAA 600
        |   || ||| ||||||||||||||||||| |||| ||||||||||||
Db  502 GGCGACAAGTGAGACTCTGTCTCAAAAAAACAAAAACAAAACCACTATA 561

Qy  601 T 601
        |
Db  562 T 562
```

```
FEATURES             Location/Qualifiers
     source          1..523
                     /organism="Homo sapiens"
                     /db_xref="GDB:7635800"
                     /db_xref="taxon:9606"
                     /clone="RPCI-11-354K9"
                     /clone_lib="RPCI-11"
                     /sex="Male"
                     /cell_type="Lymphocytes"
                     /note="Vector: pBACe3.6; Site_1: EcoRI; Site_2: EcoRI;
                     RPCI11 Human Male BAC Library"
                     Class: BAC ends.
BASE COUNT      119 a    122 c    110 g    172 t
ORIGIN
```

```
RESULT 7
AZ517953/c
LOCUS       AZ517953       523 bp    DNA     linear   GSS 16-OCT-2000
DEFINITION  RPCI-11-354K9.TVB RPCI-11 Homo sapiens genomic clone RPCI-11-354K9,
            DNA sequence.
ACCESSION   AZ517953
VERSION     AZ517953.1  GI:10827968
KEYWORDS    GSS.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 523)
  AUTHORS   Zhao,S., Adams,M.D., Nierman,W., Malek,J., de Jong,P. and Venter
            ,J.C.
  TITLE     BAC end sequences of library RPCI-11
  JOURNAL   Unpublished (1997)
COMMENT     Other_GSS: RPCI-11-354K9.TJ
            Contact: Shaying Zhao
            The Institute for Genomic Research
            Department of Eukaryotic Genomics
            9712 Medical Center Dr., Rockville, MD 20850, USA
```

Tel: 301 838 0200
Fax: 301 838 0208
Email: szhao@tigr.org
Clones are derived from the human BAC library RPCI-11. For BAC
library availability, please contact Pieter de Jong
(pieter@dejong.med.buffalo.edu). Clones may be purchased from
BACPAC Resources (http://bacpac.med.buffalo.edu/ordering) or from
Research Genet cs (info@resgen.com). BAC end search page:
http://www.tigr.org/tdb/humgen/bac_end_search/bac_end_search.html.
This BAC end was generated during the R&D process and may have
higher chance of clone tracking errors.
Seq primer: T7
Class: BAC ends.

```
RESULT 8
AZ773470/c
LOCUS       AZ773470       449 bp    DNA     linear   GSS 16-FEB-2001
DEFINITION  UP_524-15C_T7 RPCI11 Human Male BAC Library Homo sapiens genomic
            clone 524-15C, DNA sequence.
ACCESSION   AZ773470
VERSION     AZ773470.1  GI:12897868
KEYWORDS    GSS.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 449)
  AUTHORS   Cheung,V.G., Dalrymple,H.L., Narasimhan,S., Watts,J., Schuler,G.,
            Raap,A.K., Morley,M. and Bruzel,A.
  TITLE     A resource of mapped human bacterial artificial chromosome clones
  JOURNAL   Genome Res. 9 (10), 989-993 (1999)
  MEDLINE   99455100
COMMENT     Contact: Arcaro MA, Morley M, Burdick J, Cheung VG
            Department of Pediatrics
            University of Pennsylvania

Query Match     16.0%;   Score 159.8;  DB 17;   Length 523;
Best Local Similarity    76.1%;   Pred. No. 4.1e-16;
Matches 197;   Conservative  0;   Mismatches 62;   Indels 0;   Gaps 0;

Qy  353 TCATAACCTCAGGACCAGCCTCGGCCAACATAGTGAAAACCCTGTCTGTACTAAAAATACAA 412
        ||   |   ||||| |||||| |||||||  | |||||| || |||| |||||||||||||||
Db  288 TCAGGAGTTCAAGAACAGCCTGGCCAACGTGGTGAAACCCCGTCTCTACTAAAAATACAA 229

Qy  413 AAATTAGCTTGGCATGGTGGCGTGCGTGGGCACCTGTAATCCCAGCTACTCAGGGAGGCTGAGGCAG 472
        |||||||||| || |||||| | | | ||| |||||| |||||||||||| |||||||||||||
Db  228 AAATTAGCTGGGCGTGGTGGGCGGGCGCCTGTAGTCCCAGCTACTCGGGAGGCTGAGG 169

Qy  473 GAGAATCACTTGAATCCGGGAGGCAGAGGTTGCAGTGAGCTGAGATTGCAGTGCGAGCCGCCTG 532
        ||||||||||||||  |  ||||||||||||| |||||||| ||||| ||    || ||| |
Db  168 GAGAATCACTTGAACCCAGGAGGCAGAGGTTGTTGGTGAGCCGAGATCGCCCATTGCACTCC 109

Qy  533 GGCGACAGAGACAGAAATGAAACTCTGTCTCAAAAAACAAAACAAAACCCACTATA 592
        ||||||| | | |||| | ||||||||||||| ||||||| |||| ||||||||
Db  108 AGCCTGGGCGACAAGAGCGAGACTCTGTCTCTCAAAAAAACAAAAAAACCACTA 49

Qy  593 CATAAAAAATGAACAAATGAT 611
        ||| ||||||||||||| |
Db   48 AATGACATTTGAAAGCAAT 30
```

3516 Civic Center Blvd, ARC 516, Philadelphia, PA 19104, USA
Tel: 215 590 2664
Fax: 215 590 3709
Email: mlennox@mail.med.upenn.edu
Plate: 524 row: C column: 15
Seq primer: T7
Class: BAC ends.

```
FEATURES             Location/Qualifiers
     source          1. .449
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="524-15c"
                     /clone_lib="RPCI11 Human Male BAC Library"
                     /sex="Male"
                     /cell_type="Lymphocytes"
                     /note="Vector: pBACe3.6; RPCI11 Human Male BAC Library"
BASE COUNT      98 a      112 c      84 g      154 t      1 others
ORIGIN
```

RESULT 9
BM686844

```
LOCUS       BM686844      394 bp      mRNA      linear      EST 28-FEB-2002
DEFINITION  UI-E-CR1-aec-d-07-0-UI.r1 UI-E-CR1 Homo sapiens cDNA clone
            UI-E-CR1-aec-d-07-0-UI 5', mRNA sequence.
ACCESSION   BM686844
VERSION     BM686844.1  GI:19000102
KEYWORDS    EST.
SOURCE      human.
ORGANISM    Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1 (bases 1 to 394)
AUTHORS     Bonaldo,M.F., Lennon,G. and Soares,M.B.
TITLE       Normalization and subtraction: two approaches to facilitate gene
            discovery
JOURNAL     Genome Res. 6 (9), 791-806 (1996)
MEDLINE     97044477
COMMENT     Contact: Soares, MB
            Program for Rat Gene Discovery and Mapping
            University of Iowa
            451 Eckstein Medical Research Building Iowa City, IA 52242, USA
            Tel: 319 335 8250
            Fax: 319 335 9565
            Email: msoares@blue.weeg.uiowa.edu
            Tissue Procurement: Dr. Gregg Hageman
            cDNA Library Preparation: Dr. M. Bento Soares, Univeristy of Iowa
            cDNA Library Arrayed by: Dr. M. Bento Soares, Univeristy of Iowa
            DNA Sequencing by: Dr. M. Bento Soares, University of Iowa
```

```
Query Match            15.8%;   Score 158.4;   DB 17;   Length 449;
Best Local Similarity  72.9%;   Pred. No. 7.4e-16;
Matches    204;   Conservative    0;   Mismatches   76;   Indels   0;   Gaps   0;

Qy    334 AGGTGTTCTTTGAAAAAAATCATAACCTCAGGACACCCTGGCCCACACATGGTGAAAACCCT 393
          || ||| ||| ||||||||||| |||| |||| || ||||||| |||||||| ||||||||
Db    280 AGGTGGGCGGGATCATPAAGGTCAGGAGATCGAGACCATCCTGCCCAACATGGTCAAAACCCC 221

Qy    394 GTCTGTPACTAAAAATACAAAAAAATTACCTTGGCATGGTGTCGTGGGCACCTGTAATCCCAGCT 453
          ||||| |||||||||||||||||| |||| |||||||| |||||||||||| ||||| ||||||||
Db    220 GTCTCTPACTAAAAATACAAAAAATTACCCAGTGTGTGTGGGCACACACCTGTAGTCCCAGCT 161

Qy    454 ACTCAGGAGGCTGAGGCAGGAGGAGAATCACTTGAATCCCGGGAAGCAGAGAGTTGCAGTGAGCT 513
          ||||  ||||||||||||||||||| |||||||||||| ||||| | || ||| ||||||||||||
Db    160 ACTCCGGAGGCTGAGGCAGGAGAATCACTTGAACCTGGGAGGCGGAGGTTGCAGTGAGCT 101

Qy    514 GAGATTGCAGTGCAGCCTGGGCGCGACAGAGCGAGAAATGAAAACTCTGTCTCTCAAAAACAAAC 573
          ||||||||| | |||| ||| || |||||||| ||| ||||| | ||| ||||||||||||| |
Db    100 GAGATTGCACGACAGCATTCCAGTCTGGCGACAGAGTGAGACTCTPATCTCAAAAAATAAA 41

Qy    574 AAACAAAAAAAACCACTATACATATAAAAATGAACAAATGATGC 613
          || ||||||||  |||  |  ||| |||| ||| || |||
Db     40 AATAAAAAAAAGAAAAAAGAAAAAAAAAAGAATTCCGC 1
```

```
FEATURES             Location/Qualifiers
     source          1. .394
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="UI-E-CR1-aec-d-07-0-UI"
                     /clone_lib="UI-E-CR1"
                     /tissue_type="eye anterior segment"
                     /dev_stage="adult"
                     /lab_host="DH10B (Life Technologies) (T1 phage resistant)"
                     /note="Organ: eye; Vector: pT7T3-Pac (Pharmacia) with a
                     modified polylinker; Site_1: EcoR I; Site_2: Not I;
                     UI-E-CR1 is a normalized cDNA library containing the
                     following tissue(s): eye anterior segment. The library was
                     constructed according to Bonaldo, Lennon and Soares,
                     Genome Research, 6:791-806, 1996. First strand cDNA
                     synthesis was primed with an oligo-dT primer containing a
                     Not I site. Double stranded cDNA was ligated to an EcoR I
                     adaptor, digested with Not I, and cloned directionally
                     into pT7T3-Pac vector. The oligonucleotide used to prime
                     the synthesis of first-strand cDNA contains a library tag
                     sequence that is located between the Not I site and the
                     (dT)18 tail. The sequence tag for this library is
                     AATGCCCGAT. This library was created for the program, Gene
                     Discovery in the Visual System, supported by National Eye
                     Institute (NEI)."
BASE COUNT      156 a      76 c      100 g      61 t      1 others
ORIGIN
```

Clone Distribution: Researchers may obtain clones from Research
Genetics (www.resgen.com).
The following repetitive elements were found in this cDNA
sequence: 9-272, >ALU 273-346, >POLY_A#Simple_repeat

RESULT 10
AA229935/c

```
LOCUS       AA229935      390 bp      mRNA      linear      EST 21-AUG-1997
DEFINITION  nc51g10.r1 NCI_CGAP_Pr3 Homo sapiens cDNA clone IMAGE:1011714
            similar to contains Alu repetitive element ;repetitive element;
            repetitive element MER4 ; mRNA sequence.
ACCESSION   AA229935
VERSION     AA229935.1  GI:1852248
KEYWORDS    EST.
SOURCE      human.
ORGANISM    Homo sapiens
```

```
Query Match            15.7%;   Score 157;   DB 14;   Length 394;
Best Local Similarity  68.2%;   Pred. No. 1.3e-15;
Matches    217;   Conservative    0;   Mismatches   101;   Indels   0;   Gaps   0;

Qy    353 TCATAAACCTCAGGACCACCCTGGCCACAACATGGTGAAAACCCTGTCTGTTGTACTAAAATACAA 412
          ||| | |||||||| |||| ||||||||  |||||||||||||||||||||| ||||||| ||||
Db     54 TCAAGAGATTGAGACCAGCCTGGCCAACATGGTGAAAACCCTGTCTCTACTAAAAACACAA 113

Qy    413 AAAATTAGCTTGGCATGGTGGCGCGCACCTGTAATCCCAGCTACTCAGGAGGCGAGGCAG 472
          |||||||||  |||| |||||||||| |||||||| |||||||||||| ||||| ||||||
Db    114 AAAATTAGCCGGGCGTGGTGGCGCGCGCCTGTAGTCCCAGCTACTCCCAGAGGCTGAGGCAG 173

Qy    473 GAGAATCACTTGAATCCCGGAGGCAGGAGGATTGCAGTGAGCTGAGATTGCGAGCCTG 532
          ||||||| ||||| ||| ||||| |||| ||| ||||||||| |||||| | | |||||
Db    174 GAGAATCGCTTGAACCCGGGAGGGGGAGGTTGCAGTGAGCCGAGATTGTGTCTCAGTGCT 233

Qy    533 GGGCGACAGAGCACAAAATGAAAACTCTGTCTCAAAAACAAACAAAAAAAACCACTATA 592
          | | |||||||| || | ||| ||||||  ||| ||||| |||||||||||||||||
Db    234 CCAGCCTGGCCAACAGAGCGAGACTCTGCGTCTCANAAAAAAAAAAAAAAAAAACAA 293

Qy    593 CATAAAAAATGAACAAATGATGCCACAATAGCACCAGAGAGAATTTTTTATAAATACAGATTCCCA 652
          || ||||||  |  |  || |||||||||| || |||||| ||||  |   | | |||||| |||
Db    294 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAATTGGGGA 353

Qy    653 GGGCCTGCCCCAGACCTA 670
          |||| ||||||| |||
Db    354 ATTGGGGCGCCCAAGCTTA 371
```

Record (left column)

```
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1 (bases 1 to 390)
AUTHORS     NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
TITLE       National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
            Tumor Gene Index
JOURNAL     Unpublished (1997)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: W. Marston Linehan, M.D., Rodrigo Chuaqui, M.D.
            , Michael Emmert-Buck, M.D., Ph.D.
            cDNA Library Preparation: David B. Krizman, Ph.D.
            cDNA Library Arrayed by: Genome Systems Inc., Greg Lennon, Ph.D.
            DNA Sequencing by: Washington University Genome Sequencing Center
            Clone distribution: NCI-CGAP clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            www-bio.llnl.gov/bbrp/image/image.html
            Insert Length: 760    Std Error: 0.00
            Seq primer: -28m13 rev1 ET from Amersham
            High quality sequence stop: 351.
FEATURES             Location/Qualifiers
     source          1..390
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:1011714"
                     /clone_lib="NCI_CGAP_Pr3"
                     /sex="Male"
                     /dev_stage="45 years old"
                     /lab_host="DH10B"
                     /note="Vector: pAMP10; Site_1: NotI; Site_2: EcoRI; 1st
                     strand cDNA was primed with oligo(dT)17 on 50 ng of
                     DNase-treated, total cellular RNA obtained from 5,000-10
                     ,000 microdissected cells histologically-determined to be
                     fully malignant prostate cancer cells. Double-stranded
                     cDNA was ligated to EcoRI adaptors, 5 cycles of PCR
                     applied to the cDNA with an adaptor-specific primer, and
                     the resulting PCR product subcloned into pAMP10 by the
                     UDG-cloning method (Life Technologies). Average insert
                     size is 600 bp. NOTE: Not directionally cloned. This
                     library was constructed by David Krizman."
BASE COUNT       80 a     101 c      84 g     125 t
ORIGIN

Query Match         15.7%;   Score 156.8;  DB 9;   Length 390;
Best Local Similarity   78.3%;   Pred. No. 1.4e-15;
Matches  188;  Conservative   0;  Mismatches  52;  Indels   0;  Gaps   0;

Qy    350 AAATCATAACCTCAGGACCAGCCTGGCCAACATGGTGAAAACCCTGTCTTGTACTAAAAATA 409
          | |||| |||||||||||||||||||||||||||| |||||||||||||| ||||||||
Db    269 AGATCAGGAGTTCGAGACCAGCCTGGCCAACATGGCAAAAACCCTGTCTTCCTACTAAAATTA 210

Qy    410 CAAAAATTAGCTTGGCCATGGTCGTGGGCCACCTGTAATCCCAGCTACTCAGGAAGCTGAGG 469
          ||||||||||| |||| |||||| ||||| |||||| ||||||||||||| |||||||||||
Db    209 CAATAAATTAGCCAGGCATGGTGGCGGGCGCCTGTGCATCCCAGCTACTTGGGAGGCTGAGG 150

Qy    470 CAGGAGAATCACTTGAATCCGGGAGGCGGAGAGTTGCAGTGAGTCGAGTCGAGC 529
          ||||||||||| |||||| ||| ||||| ||||||||||||||| |||
Db    149 CAGGAGAATCGCTTGAACCCAGGAGGCAGTGAGTTGCAGTGAGCCACCACTGC 90

Qy    530 CTGGGCGACAGAGCAAGAAATGAAACTCGTGTCTCAAAAACAAACAAAAAAAACCACT 589
          | | |||  |||||||||| | |||  ||| |||||||||||||||||||| |||
Db    89  ATGGCAAGCCTGGGCAAATAGAGCGAGCTCCGTCTCAAAAAAAAAAAAAAAAAAGT 30
```

RESULT 11 (left column, continued)

```
RESULT 11
LOCUS       BI823844    664 bp    mRNA    linear    EST 04-OCT-2001
DEFINITION  60303006F1 NIH_MGC_115 Homo sapiens cDNA clone IMAGE:5179920 5',
            mRNA sequence.
ACCESSION   BI823844
VERSION     BI823844.1  GI:15935394
KEYWORDS    EST.
SOURCE      human.
```

Record (right column)

```
ORGANISM    Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1 (bases 1 to 664)
AUTHORS     NIH-MGC http://mgc.nci.nih.gov/.
TITLE       National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL     Unpublished (1999)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: Life Technologies, Inc.
            cDNA Library Preparation: Life Technologies, Inc.
            cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
            DNA Sequencing by: Incyte Genomics, Inc.
            Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            High quality sequence start: 3
            High quality sequence stop: 569.
FEATURES             Location/Qualifiers
     source          1..664
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:5179920"
                     /clone_lib="NIH_MGC_115"
                     /lab_host="DH10B"
                     /note="Organ: pooled brain, lung, testis; Vector:
                     pCMV-SPORT6; Site_1: NotI; Site_2: EcoRV (destroyed); RNA
                     source anonymous pool of 6 male brains, age range 23-27; 1
                     male lung, age 27; and 1 male testis, age 69. Library is
                     oligo-dT primed and directionally cloned (EcoRV site is
                     destroyed upon cloning). Average insert size 1.8 kb,
                     insert size range 1-3 kb. Library is normalized and
                     enriched for full-length clones and was constructed by C.
                     Gruber (Invitrogen). Research Genetics tracking code
                     021. Note: this is a NIH_MGC Library."
BASE COUNT      159 a    178 c    179 g    148 t
ORIGIN

Query Match         15.7%;   Score 156.6;  DB 13;  Length 664;
Best Local Similarity   60.9%;   Pred. No. 1.2e-15;
Matches  309;  Conservative   0;  Mismatches 164;  Indels  34;  Gaps   2;

Qy    1   CCTGCCCACCACATGCCTGGCCTAATTTCTTATTTTAGTAGAGACGAGGTTTGCCATGTTG 60
          |||||| ||||||  |||||| ||| ||| ||| ||||||||||| |||| ||||| |||||
Db    107 CCTGCCACCACACCTGGCTAGTTTGTTTGAACTTTTAGTAGAGACAGGGTTTGCCGTGTTG 166

Qy    61  ACCAGGCTGGTCTCGAAACTCCTTGACCTCAGGTGATCGGCCTGCCTCGCCAAAGTG 120
          || ||||| ||||||| ||||| || |||||||||||| |||  ||  |||||||||||
Db    167 GTCATGCCTTGTCTCGAACTCCTGACCTCAGGTGATCTGCCCACCTAGGCTTCCCAAAGTG 226

Qy    121 CTGGGATTATAGGCGTGAGCCGCCGCCGCCCAGCCAACATGTTTTAAATACTGAAAAGTAG 180
          ||||| ||| ||||||||||| |||| || |||| ||||
Db    227 CTGGGTTTACCGGCGTGAGCCACCGCACCTGGCCCTCAAATTTT---------------- 268

Qy    181 AGGGAATAAGTTATAGTGTACCCCATTTACCCATCACTCAGTTTCAACAGTTCAGTGACATA 240
                    |||| || |  |||| |||   ||| |||||
Db    269 ----------GTTTTCTTCTATACCAGTTGCCAGGTATCATTGAATGAAGTGGTGGTCAGC 314

Qy    241 TTTATTTCTTCTATACCAGTTACTCTCCCCACTGGGATTATTTTTTAAGGCAAAACC 300
          ||| || ||  |||||
Db    315 TTAAGGAAAAACAACTAAGAACACTA--CTTAAGAAAGATAAGAGGCCAGCGAGGTGGC 372

Qy    301 AGATGACATTTTATCCCTTAAATACTTTAGATAAAAGGTGTTCTTGAAAAAAATCATAACC 360
          |||
Db    373 TCATGCCTGTAATCCCAGCACTTTGGGAGGCCGAGGCGGGCAGATCACGAGGTCAAGAGA 432

Qy    361 TCAGGACCAGCCTGGCCAACATGGTGAAAACCCTGTCTGTGTACTAAAAATACAAAAATTAGC 420
          |||  ||| ||||||| ||||  |||||| |||| | | ||||||||||||||||| |||||
Db    433 TCAACACCATCCTGGCTAACGGGGTGAAACCCCGTCTCTACTAAAAATACAAAAGATTAGC 492

Qy    421 TTGGCATGGTCGTGGGCCACCTGTAATCCCAGCTACTCAGGAAGCTGAGGCAGGAGAATCA 480
          |||| |||||||||||||||||||
Db    ... 
```

RESULT 12

```
Qy    481 CTTGAATCCGGGAAGCAGAGATTGCAG 507
          ||||||||||||||||||||||||||||
Db    553 TGTGAACCAGGAGGCAGACCTTGCAG 579
```

```
Db    493 CAGACGTGGTGGCGGGGTGCCTGTCGTAGTCCCAGCTACTCGAGGAGGCTGAGGCAGGAGAATGG 552
```

LOCUS   AQ588276/c

DEFINITION CITBI-E1-2635H18.TR CITBI-E1 Homo sapiens genomic clone 2635H18,
            DNA sequence.
ACCESSION   AQ588276
VERSION   AQ588276.1 GI:5014956
KEYWORDS   GSS.
SOURCE   human.
  ORGANISM Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1 (bases 1 to 502)
  AUTHORS   Zhao,S., Adams,M.D., Nierman,W., Malek,J., Shizuya,H., Simon,M. and
            Venter,J.C.
  TITLE   Use of BAC End Sequences from CalTech Libraries for Sequence-Ready
            Map Building
  JOURNAL   Unpublished (1997)
COMMENT   Other_GSSs: CITBI-E1-2635H18.TF
            Contact: Shaying Zhao, William Nierman, Mark Adams
            The Institute for Genomic Research
            Department of Eukaryotic Genomics
            9712 Medical Center Dr., Rockville, MD 20850
            Tel: 301 838 0200
            Fax: 301 838 0208
            Email: hbe@tigr.org
            Clones are available from Research Genetics (info@resgen.com). BAC
            end search page:
            http://www.tigr.org/tdb/humgen/bac_end_search/bac_end_search.html.
            Seq primer: M13 Reverse
            Class: BAC ends.
FEATURES   Location/Qualifiers
     source   1..502
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="2635H18"
                     /clone_lib="CITBI-E1"
                     /sex="male"
                     /cell_type="sperm"
                     /note="Vector: pBeloBAC11; Site_1: EcoRI; Site_2: EcoRI;
                     CalTech Human BAC Library D"
BASE COUNT   147 a   115 c   100 g   140 t
ORIGIN
```

```
RESULT 12
LOCUS       CITBI-E1-2635H18.TR CITBI-E1 Homo sapiens genomic clone 2635H18,
                                   502 bp    DNA    linear    GSS 07-JUN-1999
```

```
Query Match            15.6%;  Score 156.2;  DB 17;  Length 502;
Best Local Similarity  75.5%;  Pred. No. 1.6e-15;
Matches   194;  Conservative    0;  Mismatches   63;  Indels    0;  Gaps    0;
```

```
Qy    328 AGATAAAGGTGTTCTTTGAAAAAAAAAATCATAACCTCAGACACCAGCCTGGCCAACATGGTGA 387
          ||  ||||||  | |||| ||||||||||  ||||||||||||||| |||||||||||||||||
Db    347 AGGCCAAGGTGGACGGATCATGAGAATTGAGAGAGATTGAGATAATCCTGGCCAACATGGTGA 288
```

```
Qy    388 AACCCTGTGTCTGTTACTAAAAAAATACAAAAATTAGCTTGGCATGGTGGTCGTGGGCACCTGTAATC 447
          |||||||| |||||||||||||||||| |||||||||| ||||||||||||||||||||||||||||
Db    287 AATCCTGTCTCTACTAAAAATACAAAAATTAGCTGGCATGGTGGCACGCACCTGTAGTC 228
```

```
Qy    448 CCAGCTACTCAGGGAAGCTGAGGCAGGAGAATCACTTGAATCCGGGAGGCAGCAGAGATTGCAG 507
          ||||||||| ||||||||||||||||||||| || ||||||||||| ||||||||||||||||
Db    227 CCAGCTACTCAGGGAGGCTGAGGCAGCACAATCGCTTGAACCTGGGAGGCGGAGGTTGCAG 168
```

```
Qy    508 TGAGCTGAGATTGCACTCGACCCTGGGCGACAGAGACGAAATGAAACTCTGTCTCAAAA 567
          ||||||||||||||| |||   | ||||||||||||||||||| |||||||||||||||
Db    167 TGAGCTGAGATTAGCACCACTACTCCAGCCTGGGGACAGAGAGTGAGACTCTGTCTCAAAA 108
```

```
Qy    568 ACAAACAAACAAAAAAA 584
          |||| || | |||||||
Db     57 (not shown)
```

RESULT 13

```
Db    107 AAAAAAAAGAAAAAAAA 91
```

LOCUS   AQ549290

DEFINITION RPCI-11-428K15.TV RPCI-11 Homo sapiens genomic clone RPCI-11-428K15
            , DNA sequence.
ACCESSION   AQ549290
VERSION   AQ549290.1 GI:4908467
KEYWORDS   GSS.
SOURCE   human.
  ORGANISM Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1 (bases 1 to 566)
  AUTHORS   Zhao,S., Adams,M.D., Nierman,W., Malek,J., de Jong,P. and Venter
            ,J.C.
  TITLE   Use of BAC End Sequences from Library RPCI-11 for Sequence-Ready
            Map Building
  JOURNAL   Unpublished (1997)
COMMENT   Other_GSSs: RPCI-11-428K15.TJ
            Contact: Shaying Zhao, William Nierman, Mark Adams
            The Institute for Genomic Research
            Department of Eukaryotic Genomics
            9712 Medical Center Dr., Rockville, MD 20850
            Tel: 301 838 0200
            Fax: 301 838 0208
            Email: hbe@tigr.org
            Clones are derived from the human BAC library RPCI-11. For BAC
            library availability, please contact Pieter de Jong
            (pieter@dejong.med.buffalo.edu). Clones may be purchased from
            BACPAC Resources (http://bacpac.med.buffalo.edu/ordering) or from
            Research Genet cs (info@resgen.com). BAC end search page:
            http://www.tigr.org/tdb/humgen/bac_end_search/bac_end_search.html.
            Seq primer: T7
            Class: BAC ends.
FEATURES   Location/Qualifiers
     source   1..566
                     /organism="Homo sapiens"
                     /db_xref="GDB:7664222"
                     /db_xref="taxon:9606"
                     /clone="RPCI-11-428K15"
                     /clone_lib="RPCI-11"
                     /sex="Male"
                     /cell_type="Lymphocytes"
                     /note="Vector: pBACe3.6; Site_1: EcoRI; Site_2: EcoRI;
                     RPCI11 Human Male BAC Library"
BASE COUNT   181 a   120 c   139 g   126 t
ORIGIN
```

```
RESULT 13
LOCUS       AQ549290                        566 bp    DNA    linear    GSS 28-MAY-1999
DEFINITION  RPCI-11-428K15.TV RPCI-11 Homo sapiens genomic clone RPCI-11-428K15
```

```
Query Match            15.6%;  Score 156.2;  DB 17;  Length 566;
Best Local Similarity  76.7%;  Pred. No. 1.5e-15;
Matches   191;  Conservative    0;  Mismatches   58;  Indels    0;  Gaps    0;
```

```
Qy    353 TCATAACCTTCAGGACCCAGCCTGGCCAACATGGTGAAACCCTGTCTGTACTAAAATACAA 412
          |||  ||||||| | || ||||||||||||||||||||| |||| ||||||||| |||||||
Db    137 TCAAGAAAATCAAGACCATCCTGGCCAACATGGTGAAGCCCTGTCTTCTACTAAAATACAA 196
```

```
Qy    413 AAAATTAGCTTGGCATGGTGGTCGTGGGCACCTGTAATCCCAGCTACTCTCAGGGAGGCTGAGGCAG 472
          |||||| ||||||||||||||||||||||||||||| |||||||||| |||||||||||||||||||
Db    197 AAATTAACTGGGCCGTGGTGGCAGGCACCTGTAATCCCAGCTACTCCAGGAGGCTGAGGCAG 256
```

```
Qy    473 GAGAATCACTTGAATCCGGGAAGCAGCAGAGATTGCAGTGAGCTGAGATTGCAGTGAGCCTG 532
          |||||||||||||||||||| |||| |||||||||||||||||||||||| |||||||||||
Db    257 GAGAATCACTTGAACCCAGGAGGCAGCGAGGTTGCAGTGAGCTGAGATCGCGGCCACTGCACC 316
```

```
Qy    533 GGCCGACAGAGACAGAAATGAAACTCTGTCTCAAAAACAAAACAAAACAAAAAAACCACTATA 592
          ||||  |   |||||||||||| ||||||||||||| || |||||||||||||||||||||
Db    317 CCAGCCTCGATGACGAGAGTGAGACTCTGTCAGTCTCAAAAAAAAAAAAAAAATCCCTACTTAT 376
```

```
Qy    593 CATAAAAAT 601
          |||  |
Db     ...
```

```
Db    377 GTTAAGAGT 385

RESULT 14
BM970657/c
LOCUS       BM970657          682 bp   mRNA   linear   EST 21-MAR-2002
DEFINITION  UI-CF-ECl-acj-c-09-0-UI.s1 UI-CF-ECl Homo sapiens cDNA clone
            UI-CF-ECl-acj-c-09-0-UI 3', mRNA sequence.
ACCESSION   BM970657
VERSION     BM970657.1  GI:19582244
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 682)
  AUTHORS   Bonaldo,M.F., Lennon,G. and Soares,M.B.
  TITLE     Normalization and subtraction: two approaches to facilitate gene
            discovery
  JOURNAL   Genome Res. 6 (9), 791-806 (1996)
  MEDLINE   97044477
COMMENT     Contact: McCray, PB
            McCray Lab
            University of Iowa
            2024 University of Iowa Med Labs, Iowa City, IA 52242, USA
            Tel: 319 356 4866
            Fax: 319 356 7171
            Email: paul-mccray@uiowa.edu
            Tissue Procurement: Dr. M.J. Welsh, University of Iowa
            cDNA Library preparation: Dr. M. Bento Soares, University of Iowa
            cDNA Library Arrayed by: Dr. M. Bento Soares, University of Iowa
            DNA Sequencing by: Dr. M. Bento Soares, University of Iowa
            Clone Distribution: Researchers may obtain clones from Research
            Genetics (www.resgen.com).
            The following repetitive elements were found in this cDNA
            sequence: 11-300, >ALU (matched compliment) 434-575, >ALU
            POLYA=yes.
            Seq primer: M13 FORWARD
FEATURES             Location/Qualifiers
     source          1..682
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone_lib="UI-CF-ECl"
                     /clone="UI-CF-ECl-acj-c-09-0-UI"
                     /tissue_type="Lung"
                     /dev_stage="Adult and Fetal"
                     /lab_host="DH10B (Life Technologies)"
                     /note="Organ: Lung; Vector: pT7T3-Pac (Pharmacia) with a
                     modified polylinker; Site_1: EcoR I; Site_2: Not I;
                     UI-CF-ECl is a normalized cDNA library containing the
                     following tissue(s): Normal lung from adult and from fetal
                     day 64, day 87, week 19 and week 42. The library was
                     constructed according to Bonaldo, Lennon and Soares,
                     Genome Research, 6:791-806, 1996. First strand cDNA
                     synthesis was primed with an oligo-dT primer containing a
                     Not I site. Double stranded cDNA was ligated to an EcoR I
                     adaptor, digested with Not I, and cloned directionally
                     into pT7T3-Pac vector. The oligonucleotide used to prime
                     the synthesis of first-strand cDNA contains a library tag
                     sequence that is located between the Not I site and the
                     (dT)18 tail. The sequence tag for this library is
                     AAGTGCTTAC.
                     TAG_LIB-UI-CF-ECl
                     TAG_TISSUE=Normal Lung Epithelial Cells Tissue nos 369-371
                     and 380-383
                     TAG_SEQ=AAGTGCTTAC"
BASE COUNT    145 a    172 c    177 g    185 t    3 others
ORIGIN

Query Match    15.6%;  Score 156.2;  DB 14;  Length 682;
Best Local Similarity  57.9%;  Pred. No. 1.3e-15;
Matches 338;  Conservative  0;  Mismatches 234;  Indels 12;  Gaps 3;
```

```
Qy     4 GCCACCATGCCTGGCTAAATTTCTTAT TTTTAGAGAGGAGGTTTTTGCCATGTTGACC 63
           ||||| ||||||||||| ||||||||| ||   |||| |||| |  |  ||| |||| |||
Db   577 GGCACCACCGCCTGGCTGAATTTTGTGTNTAPACTAGAGTGCGGGGATTTCGCGCGTTTGGTC 518

Qy    64 AGGCTGGTCTCGGAACTCTTGACCTCAGGTGAGTCCCGCCTGCCTCCAGCCTCCCAAAGTGCTG 123
           |||||||||||||||||||||||| |||||| ||||||||||||||||||||||||||
Db   517 AGGCTGGTCTCGGAATTCCTGGCCTCAGGCGTGTCTCTCCCCTCGGCCTCCCAAAGTGCAG 458

Qy   124 GGATTATAGGCGTGAGCCCGCGGCCACCCAGCCCAACATTTTTTAAATACTGAAAAGTAGAGG 183
           |||||| | ||                              ||| || |
Db   457 GGATTACAG----CAGGAGCCACCGTGCCCCATATATGCCATCTTTCAACTGCCTGT 402

Qy   184 GAATAGTTATTAGTGTGTACCCCATTTTACCCATCACTCAGTTTCAACAGCTGGTTGACATATTT 243
           |||| ||||| ||
Db   401 ATTTCAGAGACTCTGTCACC---TCCCTCAAGCGCCTTCGTGCGGGAGGAAAAAACTAGACAGTGAA 345

Qy   244 ATTTCTTCTATACCAGTACCTACTCCCCCACTGGGATTATTTTPAAGGGAAAAACCCAGA 303
           || |  || | ||     ||    |||   ||  ||
Db   344 ATCCAGCCAAACGAGATGAACTAACATTAAAAATTCAAAAAATTGGGCCACAGGCACGGTGGC 285

Qy   304 TGACATTTTTATCCCCTAAAATACTTTAGATAAAGGTGTTTCTTTGAAAAAAAT-----CATAA 358
           ||||| | |||  |            |||| ||   | | |||        |||| |   |||
Db   284 TCATGTTGTTGCATCCCAGCACTTTGGGAGGCCGAGGCGGGGGATCACCTGAGGTCAGGA 225

Qy   359 CCTCAGGACCAGCCTGGGCCAACATGGTGAAACCCTGTCTGTACTAAAAAATACAAAAAATTA 418
           ||||| ||||| ||||||||||||||||||| ||||| ||||||||| |||||||||| ||
Db   224 GTTTCGGACCAGCCTGGCCAACATGGCGAAACCCCCGTCTCTACTAAAAAATACAAAAAATTA 165

Qy   419 GCTTGGCATGGTGGTGGGCGCACCTGTAATCCCAGCTACTCAGGAAGCTGAGGCAGGAGGAAT 478
           |||||||||||||||| ||| |||| ||||||||||||||| || ||| |||| ||| |||
Db   164 GCTGGGCATGATGGCGGGCGCCTGTAATCCCAGCTACTTGGGAGGCTATGGCAGTGAAAAT 105

Qy   479 CACTTGAATCCGGAAGCAGAGATTGCAGTGAGCTGAGATTGCAGTGCAGTGCGAGCCTGGGCGAC 538
           |||| ||| ||| |||| |||| ||| ||||||| |||| |  ||  |||| |||  |||
Db   104 CGCTTGAACCTGGGAGGCAGAAGTTGCAGTGAGCCAAGACCACGCCCGTTGCACTCCAGCC 45

Qy   539 AGAGACAAAATGAAAACTCTGTCTCAAAAACAAACAAACAAAAAAAAA 582
           ||| | |  ||  || |||||  ||||||| |||| ||| ||||||||
Db    44 TGGGCAACAAGAGCAAAAGTCATCTCAAAAAAAAAAAAAAAAAAAAAA 1
```

```
RESULT 15
BF678427
LOCUS       BF678427          702 bp   mRNA   linear   EST 21-DEC-2000
DEFINITION  602085862F1 NIH_MGC_83 Homo sapiens cDNA clone IMAGE:4250246 5',
            mRNA sequence.
ACCESSION   BF678427
VERSION     BF678427.1  GI:11952322
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 702)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: CLONETECH Laboratories, Inc.
            cDNA Library Preparation: CLONETECH Laboratories, Inc.
            cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
            DNA Sequencing by: Incyte Genomics, Inc.
            Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLCM1071 row: k  column: 15
            High quality sequence stop: 562.
FEATURES             Location/Qualifiers
     source          1..702
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:4250246"
```

/clone_lib="NIH_MGC_83"
/lab_host="DH10B (T1 phage-resistant)"
/note="Organ: prostate; Vector: pDNR-LIB (Clontech);
Site_1: SfiI (ggccgcctcggcc); Site_2: SfiI (ggccattatggc
); 5' and 3' adaptors were used in cloning as follows: 5'
adaptor sequence: 5'-CACGGCCATTATGGCC-3' and 3' adaptor
sequence: 5'-ATTCTAGAGGCCGAGGCGGCCGACATG-dT(30)BN-3'
(where B = A, C, or G and N = A, C, G, or T). Average
insert size 1.4 kb (range 0.5-4.0 kb). 14/15 colonies
contained inserts by PCR. This library was enriched for
full-length clones and was constructed by Clontech
Laboratories (Palo Alto, CA)."

BASE COUNT   223 a   170 c   210 g   99 t
ORIGIN

Query Match   15.6%;   Score 156;   DB 12;   Length 702;
Best Local Similarity   79.2%;   Pred. No. 1.4e-15;
Matches 198;   Conservative   0;   Mismatches   50;   Indels   2;   Gaps   1;

Qy   353 TCATAACCTCAGGACCAGCCTGGCCAACACATGGTGAAACCCTGTCTGTACTAAAAAATACAA 412
          ||| ||||||||||||||| ||||||||||||||||||||||||||||||| ||||||||||||
Db   344 TCAGGAGTTCGAGGCCAGCCTGGTCAACACATGGTGAAACCCTGTCTCTACTAAAAAATACAA 403

Qy   413 AAATTAGCTTGGCATGGTCGTTGGGGCACCTGTAATCCCAGCTACTCAGGAAGCTGAGGCAG 472
          ||||||||| ||||||||||||| |||||||||||||||||||| |||||||||||||||||||
Db   404 AAATTAGCTGGGCATGGTGGCGCACCTGTAATCCCGGCTACTCGGGAGGCTGAGGCAG 463

Qy   473 GAGAATCACTTGAATCCGGGAAGCAGAGAATTGCAGTGAGCTGCAGTTCGAGCCTG 532
          ||||||||| |||||| |||||||| ||| |||||||||| ||| ||||||||||
Db   464 GAGAATCATTTGAACCCGGGAGGCGGAGGCTGCAGTGAGCCGAGATTGCGCCATAGCACT 523

Qy   533 GGGCGACGAG--ACAGAAATGAAACTCTGTCTCAAAAACAAAACAAAAAAAAAACCACTA 590
          ||| | | || || |||||||||||||||||||||||||||||| |||||||| |
Db   524 CCAGACTGGGCAACACAGGGTGAGACTCTGTCTCAAAAAAAAAGAAAACAGAAAAAAAAAA 583

Qy   591 TACATAAAAA 600
          ||| ||| |||
Db   584 AACAAACAAA 593

Search completed: March 29, 2003, 22:20:49
Job time : 1848.16 secs

GenCore version 5.1.4_p5_4578
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 29, 2003, 16:26:54 ; Search time 46.4345 Seconds
                                            (without alignments)
                                            6604.503 Million cell updates/sec

Title: US-09-691-219-3_COPY_1_1000
Perfect score: 1000
Sequence: 1 cctgccaccatgcctggcta...........caagaggctgcacctctga 1000

Scoring table: IDENTITY_NUC
                 Gapop 10.0 , Gapext 1.0

Searched: 441362 seqs, 153338381 residues

Total number of hits satisfying chosen parameters: 882724

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
                 Maximum Match 100%
                 Listing first 45 summaries

Database :   Issued_Patents_NA:*
                 1: /cgn2_6/ptodata/1/ina/5A_COMB.seq:*
                 2: /cgn2_6/ptodata/1/ina/5B_COMB.seq:*
                 3: /cgn2_6/ptodata/1/ina/6A_COMB.seq:*
                 4: /cgn2_6/ptodata/1/ina/6B_COMB.seq:*
                 5: /cgn2_6/ptodata/1/ina/PCTUS_COMB.seq:*
                 6: /cgn2_6/ptodata/1/ina/backfiles1.seq:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

	Result NO.	Score	% Query Match	Length	DB	ID	Description
	1	217.6	21.8	59065	4	US-09-813-817-3	Sequence 3, Appli
	2	217.6	21.8	59065	4	US-09-978-197-3	Sequence 3, Appli
	3	197.8	19.8	70000	4	US-09-851-896-3	Sequence 3, Appli
	4	197.2	19.7	12394	4	US-09-488-856A-10	Sequence 10, Appli
	5	197.2	19.7	35060	3	US-08-814-095-7	Sequence 7, Appli
	6	189.8	19.0	162450	3	US-09-345-882-1	Sequence 1, Appli
c	7	189	18.9	35060	3	US-08-814-095-7	Sequence 7, Appli
c	8	186.4	18.6	17327	1	US-07-906-871-15	Sequence 15, Appli
	9	181	18.1	162450	4	US-09-345-882-1	Sequence 1, Appli
c	10	180.6	18.1	16389	4	US-09-741-154-3	Sequence 3, Appli
c	11	174.8	17.5	59065	4	US-09-813-817-3	Sequence 3, Appli
c	12	174.8	17.5	59065	4	US-09-978-197-3	Sequence 3, Appli
c	13	171.4	17.1	72604	4	US-09-268-992-7	Sequence 7, Appli
c	14	171.4	17.1	72604	4	US-09-657-474-7	Sequence 7, Appli
c	15	169.6	17.0	99500	4	US-09-798-096-10	Sequence 10, Appl
c	16	168.8	16.9	84495	4	US-09-797-906-3	Sequence 3, Appli
c	17	166.8	16.7	16389	4	US-09-741-154-3	Sequence 3, Appli
c	18	162.4	16.2	152331	3	US-09-810-671-3	Sequence 3, Appli
c	19	161.6	16.2	152331	3	US-09-810-671-3	Sequence 3, Appli
c	20	161.2	16.1	176373	3	US-09-128-155-17	Sequence 17, Appl
c	21	161.2	16.1	21234	4	US-09-128-155-16	Sequence 16, Appl
c	22	158.4	15.8	43950	4	US-09-735-934A-3	Sequence 3, Appli
c	23	158.4	15.8	246240	2	US-08-724-394A-20	Sequence 20, Appl
c	24	158.4	15.8	246240	2	US-08-724-394A-21	Sequence 21, Appl
c	25	158.4	15.8	246240	2	US-08-724-394A-22	Sequence 22, Appl
c	26	156.6	15.7	43950	4	US-09-735-934A-3	Sequence 3, Appli
c	27	150.8	15.1	81001	4	US-09-750-580-1	Sequence 1, Appli
c	28	150	15.0	29629	4	US-09-729-995-3	Sequence 3, Appli
c	29	149.8	15.0	4326	2	US-08-852-807-12	Sequence 12, Appli
c	30	149.8	15.0	13674	4	US-08-852-807-1	Sequence 1, Appli
c	31	148.6	14.9	280	2	US-08-849-701-7	Sequence 7, Appli
c	32	147.4	14.7	713	4	US-08-943-607-24	Sequence 24, Appli
c	33	146.8	14.7	713	4	US-08-943-607-23	Sequence 23, Appli
c	34	146.8	14.7	5789	4	US-09-242-948-3	Sequence 3, Appli
c	35	146.8	14.7	28720	4	US-09-341-587-7	Sequence 7, Appli
c	36	146.2	14.6	8453	4	US-09-167-681-45	Sequence 45, Appli
c	37	146	14.6	17041	1	US-08-076-011-1	Sequence 1, Appli
c	38	145.8	14.6	713	4	US-08-943-607-25	Sequence 25, Appli
c	39	145.8	14.6	713	4	US-08-943-607-26	Sequence 26, Appli
c	40	145.6	14.6	4326	2	US-08-852-807-12	Sequence 12, Appli
c	41	145.6	14.6	13674	4	US-08-852-807-1	Sequence 1, Appli
c	42	145.6	14.6	246240	2	US-08-724-394A-20	Sequence 20, Appl
c	43	145.6	14.6	246240	2	US-08-724-394A-21	Sequence 21, Appl
c	44	145.6	14.6	246240	2	US-08-724-394A-22	Sequence 22, Appl
c	45	145.4	14.5	12394	4	US-09-488-856A-10	Sequence 10, Appli

ALIGNMENTS

RESULT 1
US-09-813-817-3
Sequence 3, Application US/09813817
Patent No. 6340583

```
;   GENERAL INFORMATION:
;   APPLICANT: YAN, Chunhua et al.
;   TITLE OF INVENTION: ISOLATED HUMAN KINASE PROTEINS, NUCLEIC
;   TITLE OF INVENTION: ACID MOLECULES ENCODING HUMAN KINASE PROTEINS, AND USES
;   TITLE OF INVENTION: THEREOF
;   FILE REFERENCE: CL001178
;   CURRENT APPLICATION NUMBER: US/09/813,817
;   CURRENT FILING DATE: 2001-03-22
;   NUMBER OF SEQ ID NOS: 4
;   SOFTWARE: FastSEQ for Windows Version 4.0
;   SEQ ID NO 3
;   LENGTH: 59065
;   TYPE: DNA
;   ORGANISM: Human
;   US-09-813-817-3

Query Match       21.8%;  Score 217.6;  DB 4;  Length 59065;
Best Local Similarity   63.2%;  Pred. No. 4.5e-48;
Matches  373;  Conservative   0;  Mismatches  204;  Indels  13;  Gaps  2;

Qy       1 CCTGCCACCATGCCTGGCTAATTTTCTTATTTTTAGTAGAGACGAGGTTTTGCCATGTTG 60
           || |||| |||| || |||||| | | |||||||||||| ||||||||||||||||||||
Db   11786 CCGGCCCCCATACCCAGCGTAATTTTTTATAATTTTTAGTAGAGAAGATGGGGTTTTGCCATGTTG 11845

Qy      61 ACCAGGCTGGTCTCGAAACTCTTGACCTCAGGTGATCGGCCTGCCTCGAGCCTCCCAAAGTG 120
           ||||||||| | ||| | |||||||||||||||| ||| || |||||||||||||||||
Db   11846 GCCAACCTAGGTCTCGAACTCCTGACCTCAAGTGATCCACCTGCCTCGGCCTCCCAAAGTG 11905

Qy     121 CTGGGATTATAGGCGTGAGCCGCGCCCACCCAGC------CAACATTTTTTAAATACTGAA 174
           |||||||| |||| |||||||| | ||||||||||     | |||||| ||||| |||||
Db   11906 CTGGGATTACAGGTGTGAGCCACTGCGCCCGGCCTGGAAGGAGTGATCTTAAAAGAAAAAAA 11965

Qy     175 AAGTAGAGGGGAATAGTTGTTATACTGTGTACCCATTTACCCACATTCAGTTGTTCAACAGCTGGT 234
           ||||| | |||||| | ||| |||| |||| |||| ||| ||| |||||||||||||
Db   11966 AAACAAAAAAAAAAACTTGACTGTGTCACTCTGGTTGTGTCTCCTACCTTGTTATACTTCCA 12025

Qy     235 GACATATTTTATTTCTTCTATACCAGTACCGTACCTCTCCCCACTGGGATTATTTTAAGGCA 294
           ||| |||| | | |||  | ||| ||||| ||||||||||| ||||||||  |||||||||
Db   12026 CAACTTCCCACGTGTTCTTTGGATAAAGAACCAAAAATCCTTACTTGGCCAGGCGGCGGTGGCT 12085

Qy     295 AAACCCAGAGATGACATTTTATCCCTAAATACTTTAGATAAAGGTGTTGTTCTTTGAAAAAATC 354
           ||||||| | ||| |||  |  | || ||  ||| || ||||| |||| ||||||||||
Db   12086 CACACCTATCATCT------CAGCACTTTGGGAGGCCGAGGCAGCAGATCATGAAGTC 12138

Qy     355 ATAACCTCAGGACCAGCCTGCCACACATGGTGAAACCCGTCTGTGTACTAAAAATACAAA 414
           | |  ||||||| ||||||||||||||||| |||||||||||||||||||||||||||||
```

RESULT 2
US-09-978-197-3

Sequence 3, Application US/09978197
Patent No. 6403353
GENERAL INFORMATION:
APPLICANT: YAN, Chunhua et al.
TITLE OF INVENTION: ISOLATED HUMAN KINASE PROTEINS, NUCLEIC
TITLE OF INVENTION: ACID MOLECULES ENCODING HUMAN KINASE PROTEINS, AND USES
TITLE OF INVENTION: THEREOF
FILE REFERENCE: CL001178DIV
CURRENT APPLICATION NUMBER: US/09/978,197
CURRENT FILING DATE: 2001-10-17
PRIOR APPLICATION NUMBER: 09/813,817
PRIOR FILING DATE: 2001-03-22
NUMBER OF SEQ ID NOS: 4
SOFTWARE: FastSEQ for Windows Version 4.0
SEQ ID NO 3
LENGTH: 59065
TYPE: DNA
ORGANISM: Human
US-09-978-197-3

Query Match   21.8%;   Score 217.6; DB 4;   Length 59065;
Best Local Similarity   63.2%; DB 4;
Matches 373; Conservative   0; Mismatches 204; Indels   13; Gaps   2;

```
Qy       1 CCTGCCACCANGCCTGGCCTAATTTCTTATTTTTAGTGAGAGACGAGGTTTTGCCATGTTG 60
           |||| ||| || ||| ||| ||  || | || |||||  ||||| || ||| | ||| ||
Db   11786 CCGGCCCCCAPACCAGCTAATTTTTATATTTTTAGTAGAGATGGGGTTTTGCCATGTTG 11845

Qy      61 ACCAGGCTGGTCTCGAACTCTTGACCTCAGGTGATCCGCCTGCCTCCCAAAGTG 120
           | || ||||| | || ||| ||| ||| | ||||| || ||| ||| |||||||
Db   11846 GCCAAGCTAGTCTCTGAAACTCCTGACCTCAAGTGATCCACCTGCCTCGGCCTCCCAAAGTC 11905

Qy     121 CTGGGATTATAGGCGTGAGCCGCCGCGACCCAGC------CAACATTTTTAAATACTTGAA 174
           |||||||| |||||| ||| ||| ||| ||||||      ||||||||| |||| |||||
Db   11906 CTGGGATTACAGGCGTGAGCCACTGCACCTCTGTTGTCTCGGCCCTCCCAAAGTG 11965

Qy     175 AAGTAGAGGGGAATAGTTATAGTGTACCCCATTCACCTAGTTCAAACACACTGGT 234
           ||||| |||| |  || ||| ||| ||| || |||| |||||| || ||| ||||
Db   11966 AAACAAAAAAAACTTGACTGTGTCACTCTCTGTTGTCTCCTACCTTGTATACTTCCA 12025

Qy     235 GACATATTTATTTCTTCTATACCAGTACCCGTACTCTCCCCACTGGGATTATTTTAAGGCA 294
           ||  |||| ||| |||  ||||| || || || ||||||| ||| || |||| ||||||
Db   12026 CAACTTCCCAGTGTTCTTTGGAATAAAGGTGTTGTTCTTTTGAAAAAATC 12085

Qy     295 AAACCCAGATGACATTTTATCCCTAAATACTTTAGATAAAGGTGTTGTTCTTTTGAAAAAATC 354
           || ||| |  | || || | ||  ||||| |||  ||||| ||| ||| ||||||||||||
Db   12086 CACACCTATCATCT-----CAGCACTTGGGAGCCGAGGCCAGATCATGAAGTC 12138

Qy     355 ATAACCTCAGGACCAGCCTGCCCAACACTGTGAAAACCCTGTCTGTTGTACTAAAAATACAAA 414
           | ||||||||| |||| ||||||||| | | ||||||| ||| ||| |||| |||||||||
Db   12139 AAGAGATTGAGACCATCCTGGCCAACATGGTGAAACCCCATCTCTACTAAAAATACAAAA 12198

Qy     415 ATTAGCTTGGCATGGTGGTGGGCGCCACCTGTAGTCCCAGCTACTTGGGAGGCTGAGGCAGGA 474
           |||||| |||| ||| |||| |||||||||| |||||||||||||| ||||| |||||||||
Db   12199 ATTAGCTGGTCGTGGTGGTGGGCGCCTGTAGTCCCAGCTACTTGGGAGGCTGAGGCAGGA 12258

Qy     475 GAATCACTTGAATCCCGGGAAGCAGAGGATTGCAGTGAGCCGAGATTGCAGTGAGCCGAGGCCTGGG 534
           |||| |||||||||||||||| |||||||||||| ||| |||||||| ||| || ||||||||
Db   12259 GAATCACTTGAACCTGGGAGGCGGAGGTTGCAGTGAGCCGAGATCACGCCACTGCACTCC 12318

Qy     535 CGACAGAGACAGAAATGAAACTCTGTCTCAAAAAACAAACAAACAAACAAAA 584
           | || ||| ||| ||| ||||| |||||||||||||||||||||| |||||||
Db   12319 AGCCTGGTGACAGAGTAAGACTCCATCTCAAAAAAAAAAAAAAAAAAAAAA 12368
```

RESULT 3
US-09-851-896-3

Sequence 3, Application US/09851896
Patent No. 6410325
GENERAL INFORMATION:
APPLICANT: C. Frank Bennett
APPLICANT: Susan M. Freier
APPLICANT: Andrew T. Watt
TITLE OF INVENTION: ANTISENSE MODULATION OF PHOSPHOLIPASE A2, GROUP VI (CA2+-INDEP
TITLE OF INVENTION: EXPRESSION
FILE REFERENCE: RTS-0220
CURRENT APPLICATION NUMBER: US/09/851,896
CURRENT FILING DATE: 2001-05-08
NUMBER OF SEQ ID NOS: 89
SEQ ID NO 3
LENGTH: 70000
TYPE: DNA
ORGANISM: Homo sapiens
FEATURE:
US-09-851-896-3

Query Match   19.8%;   Score 197.8; DB 4;   Length 70000;
Best Local Similarity   63.6%; DB 4;
Matches 396; Conservative   0; Mismatches 197; Indels   30; Gaps   5;

```
Qy       4 GCCACCANGCCTGGCTAATTTTCTTATTTTTTAGTGAGAGACGAGGTTTTGCCATGTTGACC 63
           |||||||| ||||| ||| || ||||| | || || ||||| ||||| || ||||||||||||
Db   49851 GCCACCACACCTGGCCTAATTTT-GTATTTTTTAGTAGAGACGGGGTTTCTCCATGTTGGTC 49909

Qy      64 AGGCTGGTCTCGAACTCTTGACCTCAGGTGATCCGCCTGCCTCCCAAAGTGCTG 123
           ||||||||||| ||| ||| ||| || ||||||| |||||||| ||||||||||
Db   49910 AGGCTGGTCTTGAACTCCTGACCTCAGGTGATCGGCCTGCCTCGGCCTCCCAAACTGCTG 49969

Qy     124 GGATTATAGGCGTGAGCCGCCGCCGCCCAGCCAACATTTTTAAATACTGAAAGTAGAGG 183
           |||| ||||||||||||| ||| || ||| || ||||||||| |||| |||| ||||||
Db   49970 GGATTACAAGGCGTGAGCCACCGCGCCTGGCTACTTTAAACTG------ 50018

Qy     184 GAATAGTTATAGTGTACCCCATTCACCTAGTTCAAACACTGAA 243
           ||||| |||| ||| ||| || ||||||||| |||| |||  |
Db   50019 ----ACTTAAATTATTTGAGAGAAATGTTGCATCATCTACCATAAAATGGCCATCCATCC 50074

Qy     244 ATTTCTCTATACCAGTACCGTACTCTCCCCACTGGGATTATTTTTGACGAAAAACCCAGA 303
           ||  ||| ||||| || || |||  ||| ||||| || |||||| |||| || ||||||
Db   50075 ATCCAGTGCCAACCATAAAAGGTAAC---------CAATCGCAGGGCGCGGGTGGGCTCA 50122

Qy     304 TGACATTTTATCCCTAAATACTTTAGATAAAGGTGTTCTTTGAAAAAAATCATAAACCTCA 363
           |||| ||||| ||||  ||||||| |||||| || || ||| |||| |||| ||  | |
Db   50123 CGCCTGTAATCCCAGCACATTTGGGAGGCCGAGGCGGGAATCATAAACCTCA 50182

Qy     364 GGACCAGCCTGGCCAACATGGTGAAACCCTGTCTGTACTAAAAATAC-AAAAATTAGCTT 422
           ||||||||||||||||||||| || |||||||||||| |||||||| ||||||||||||
Db   50183 AGACCAGCCTGGCCAACATGGTGAAACCCTGTCTCTACTAAAAATACAAAAAATTAGCTG 50242

Qy     423 GGCATGGTGCGTGGGCGCACCTGTAATCCCAGCTACTCAGGAAGCTGAGGCAGGAGAATCACT 482
           |||| ||| ||| |||||| |||||||||||||||| |||| || |||||||||| ||||||
Db   50243 GGCGTAGTGGTGGGCACCTGTAATCCCAGCTACTCTGGAGGCTGAGGCAGGAGAATTGCT 50302

Qy     483 TGAATCCGGGAAGCAGAGATTGCAGTGAGCTGAGATTGCAGTGAGCCGAGATCGGGGGACAGAG 542
           |||| | |||||| |||||||||||||||||| ||||| |||||| || ||| |||| || |
Db   50303 TGAACTCCGGTAGGTGAGGTTGCCAGTGAGCTGAGATCACGCCACTGCACTCCAGCCTGGG 50362

Qy     543 ACAGA-AATGAAACTCTGTCTCAAAAAACAAACAAACAAAAACCACTATACATAAAAAT 601
           |  || ||| ||||| ||| ||||||||||||||||| | ||  || | |   |||||
Db   50363 TGACAGAGTGAGACTCCGTTCTCAAAAAAAAAAAACAGTAACAACAACAAAAAAT 50422
```

```
Qy    602 GAACAAATGATGCCAACATAGCAC 624
          |||| ||  |||||   ||
Db  50423 CACAGTTAATGCTGAAAACACAC 50445
```

RESULT 4
US-09-488-8856A-10
Sequence 10, Application US/0948856A
Patent No. 6316259
GENERAL INFORMATION:
; APPLICANT:
; APPLICANT: Brett P. Monia
; APPLICANT: Robert McKay
; APPLICANT: Madeline M. Butler
; APPLICANT: Jacqueline Wyatt
; TITLE OF INVENTION: ANTISENSE MODULATION OF GLYCOGEN SYNTHASE KINASE 3 ALPHA EXP
; FILE REFERENCE: RTS-0115
; CURRENT APPLICATION NUMBER: US/09/488,856A
; CURRENT FILING DATE: 2000-01-21
; NUMBER OF SEQ ID NOS: 88
; SEQ ID NO 10
; LENGTH: 12394
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1115)...(397)
; NAME/KEY: CDS
; LOCATION: (2438)...(2625)
; NAME/KEY: CDS
; LOCATION: (5639)...(5722)
; NAME/KEY: CDS
; LOCATION: (5864)...(5974)
; NAME/KEY: CDS
; LOCATION: (7902)...(8032)
; NAME/KEY: CDS
; LOCATION: (8121)...(8227)
; NAME/KEY: CDS
; LOCATION: (9197)...(9294)
; NAME/KEY: CDS
; LOCATION: (9375)...(9470)
; NAME/KEY: CDS
; LOCATION: (9998)...(10084)
; NAME/KEY: CDS
; LOCATION: (10431)...(10523)
; NAME/KEY: CDS
; LOCATION: (11713)...(11786)
US-09-488-8856A-10

Query Match   19.7%; Score 197.2; DB 4; Length 12394;
Best Local Similarity 63.2%; Pred. NO. 6.7e-43;
Matches 373; Conservative   0; Mismatches 203; Indels 14; Gaps 4;

```
Qy    1 CCTGCCACCATGCCTGGCTAATTTCT---TATTTTTAGTAGAGACGAGGTTTGCCATG 57
        |||||||||||||||||||||||||||  ||||||||||||||||||||||| || ||
Db 4830 CCTGCCACCATGCTTGGCTAATTTTTTTGTATTTTTAGTAGAGACAGGGTTTCACCGTG 4889

Qy   58 TTGACCAGGCTGGTCTCGAAACTCTTGACCTCAGGTGATCCGCCTGCCTCCCAAA 117
        |||||||||||| ||||||||||| |||| |||||||||||||| |||||||||
Db 4890 TTAACCAGGATGGTCTCAAATCTCCTGACTTC--GTGATCCACCCACCTGCCCCAA 4947

Qy  118 GTGCTGGGGATTATAGGCGTGAGCCGCCGCCCGCCACCCAGCC--AACATTTTTTAAATACTGAA 174
        |||||||||||||||||||||||||||||||||||||||||||   ||| ||||| ||||| |||
Db 4948 GTGCTGGGATTACAGGCGTGAGCCACCACGCCCTGGCCTATCCTTTTTTAAACTT 5007

Qy  175 AAGTAGAGGGAATAGTTATAGTGTACCCCATTTACCCACTTCAACAGCTGGT 234
        || ||| ||| || ||| || |||| |  |||  |  |||||| ||||||||
Db 5008 TATTTGGAGAAAAAAATCAGAGAAGGTGCCATTGCCATTAGTTGA 5067

Qy  235 GACATATTTATTTCTTCTAT-----ACCAGTACCGTACTCTCCCCACTGGGATTATTTT 288
        | || | || |||||||||      |||||||||||||||||||||||||||||||||||
Db 5068 AAAAAAATTTTTTTTAAGTGGGTGGCTGGGCGGGTGCCTCACGCCTGTAATCCCAGC 5127
```

RESULT 5
US-08-814-095-7
Sequence 7, Application US/08814095
Patent No. 6025183
; GENERAL INFORMATION:
; APPLICANT: Soreq, Hermona
; APPLICANT: Zakut, Haim
; APPLICANT: Shani, Moshe
; TITLE OF INVENTION: TRANSGENIC ANIMAL ASSAY SYSTEM FOR
; TITLE OF INVENTION: ANTI-CHOLINESTERASE SUBSTNACES
; NUMBER OF SEQUENCES: 7
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: KOHN & ASSOCIATES
; STREET: 30500 NO. 6025183thwestern Highway, Suite 410
; CITY: Farmington Hills
; STATE: Michigan
; COUNTRY: U.S.
; ZIP: 48334
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/814,095
; FILING DATE:
; CLASSIFICATION: 800
; ATTORNEY/AGENT INFORMATION:
; NAME: Montgomery, Ilene N.
; REGISTRATION NUMBER: 38,972
; REFERENCE/DOCKET NUMBER: 2391.00066
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (248) 539-5050
; TELEFAX: (248) 539-5055
; INFORMATION FOR SEQ ID NO: 7:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 35060 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: other nucleic acid
; DESCRIPTION: /desc = "Cosmid including ACHE
; DESCRIPTION: promotor, ACHE gene and ARS gene"
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; ORIGINAL SOURCE:
; ORGANISM: Homo sapiens
; POSITION IN GENOME:
; CHROMOSOME/SEGMENT: 7q22
; FEATURE:
; NAME/KEY: promoter
; LOCATION: 4089..22464

```
Qy  289 AAGGCAAAACCCAGAGATGACACATTTTATCCCTAAATATACTTTAGATAAAGGTGTTGTTCTTTGAAA 348
        |  |||  |    | | | |||| |  || |   |||||||   || || |||||| ||   |||||
Db 5128 ACTTTGGGAGGCCTGAGGCGCTGGGATCATGAGGTCAGGAGACGGCTGAGGCAGGGGGATCAC 5187

Qy  349 AAAATCATAACCTCAGGAACCCAGCCTCGGCCCAACATGGTGGAAACCCTGTCTGTACTAAAAAT 408
        || ||| ||||| ||||||| |||| |  |  |||| |||||||| ||||||| |||||  ||
Db 5248 ACAAAAAATTAGCTGGGCGTGGTGGTGCCATGCCTGTAATCCCAGCTACTTGGGAAGGCTGAG 5307

Qy  409 ACAAAAATTAGCTTGGCATGGTGCGTGCGCCACCTGTAATCCCAGCTACTCAGGAGGCTGAG 468
        ||||||||||||| ||  | |||| |||||  |||||||||||||||||| ||  || |||||
Db 5308 GCAGGAGAATTGCTTGACCCACGGAGGCCAGAGGTTGCAGTGAGCTGAGATTGCAGTGCG 5367

Qy  469 GCAGGAGAATCACTTGAATCCGGGAGGCAGAGGATTGCAGTGAGCTGAGATTGCAGTGCGAG 528
        ||||||||||  |||  | |||||||| ||||| |||||||| | |||||||| | |||||||
Db 5368 CATGCCAGCCTGGCAACAGAGCGAAGACTCTGTCTCAAAAAAAAAAAAAAAAA 5417

Qy  529 CCTGGGCGCACGAGACGAGAAATGAAACTCTGTCTCAAAAACAAAACAAACA 578
        ||| ||  | | | | |||| |||||||||||||||||| ||| ||| |
Db 5417
```

```
OTHER INFORMATION: /function= "ACHE Promotor"
OTHER INFORMATION: /standard_name= "ACHE Promotor"
FEATURE:
NAME/KEY: exon
LOCATION: 22465..22537
OTHER INFORMATION: /function= "non-translated"
OTHER INFORMATION: /gene= "ACHE"
OTHER INFORMATION: /number= 1
FEATURE:
NAME/KEY: exon
LOCATION: 24090..25177
IDENTIFICATION METHOD: experimental
OTHER INFORMATION: /function= "(translation start:
OTHER INFORMATION: 24110)"
OTHER INFORMATION: /gene= "ACHE"
OTHER INFORMATION: /evidence= EXPERIMENTAL
OTHER INFORMATION: /gene= "ACHE"
OTHER INFORMATION: /number= 2
FEATURE:
NAME/KEY: exon
LOCATION: 25524..26009
IDENTIFICATION METHOD: experimental
OTHER INFORMATION: /evidence= EXPERIMENTAL
OTHER INFORMATION: /gene= "ACHE"
OTHER INFORMATION: /number= 3
FEATURE:
NAME/KEY: exon
LOCATION: 27005..27274
IDENTIFICATION METHOD: experimental
OTHER INFORMATION: /evidence= EXPERIMENTAL
OTHER INFORMATION: /gene= "ACHE"
OTHER INFORMATION: /number= 4
FEATURE:
NAME/KEY: exon
LOCATION: 27255..28007
IDENTIFICATION METHOD: experimental
OTHER INFORMATION: /evidence= EXPERIMENTAL
OTHER INFORMATION: /gene= "ACHE"
OTHER INFORMATION: /number= 5
FEATURE:
NAME/KEY: terminator
LOCATION: 27385..27387
FEATURE:
NAME/KEY: exon
LOCATION: 28008..28129
IDENTIFICATION METHOD: experimental
OTHER INFORMATION: /function= gene
OTHER INFORMATION: /evidence= EXPERIMENTAL
OTHER INFORMATION: /gene= "ACHE"
OTHER INFORMATION: /number= 6
FEATURE:
NAME/KEY: terminator
LOCATION: 28129..28131
FEATURE:
NAME/KEY: exon
LOCATION: complement (33779..33963)
OTHER INFORMATION: /gene= "AR"
OTHER INFORMATION: /number= 3
FEATURE:
NAME/KEY: exon
LOCATION: complement (34528..34895)
IDENTIFICATION METHOD: experimental
OTHER INFORMATION: /function= "arsenite resistance
OTHER INFORMATION: /gene= "AR"
OTHER INFORMATION: /number= 1
FEATURE:
NAME/KEY: exon
LOCATION: complement (34092..34358)
OTHER INFORMATION: /gene= "AR"
OTHER INFORMATION: /number= 2
FEATURE:
NAME/KEY: exon
LOCATION: complement (33493..33591)
OTHER INFORMATION: /gene= "AR"
OTHER INFORMATION: /number= 4

FEATURE:
NAME/KEY: exon
LOCATION: complement (33297..33408)
OTHER INFORMATION: /gene= "AR"
OTHER INFORMATION: /number= 5
FEATURE:
NAME/KEY: exon
LOCATION: complement (32959..33094)
OTHER INFORMATION: /gene= "AR"
OTHER INFORMATION: /number= 6
FEATURE:
NAME/KEY: exon
LOCATION: complement (32569..32628)
OTHER INFORMATION: /gene= "AR"
OTHER INFORMATION: /number= 7
FEATURE:
NAME/KEY: exon
LOCATION: complement (32386..32468)
OTHER INFORMATION: /gene= "AR"
OTHER INFORMATION: /number= 8
FEATURE:
NAME/KEY: exon
LOCATION: complement (31894..32080)
OTHER INFORMATION: /gene= "AR"
OTHER INFORMATION: /number= 9
FEATURE:
NAME/KEY: exon
LOCATION: complement (31363..31534)
OTHER INFORMATION: /gene= "AR"
OTHER INFORMATION: /number= 10
FEATURE:
NAME/KEY: exon
LOCATION: complement (31131..31284)
OTHER INFORMATION: /gene= "AR"
OTHER INFORMATION: /number= 11
FEATURE:
NAME/KEY: exon
LOCATION: complement (30816..31011)
OTHER INFORMATION: /gene= "AR"
OTHER INFORMATION: /number= 12
FEATURE:
NAME/KEY: exon
LOCATION: complement (30470..30626)
OTHER INFORMATION: /gene= "AR"
OTHER INFORMATION: /number= 13
FEATURE:
NAME/KEY: exon
LOCATION: complement (30187..30274)
OTHER INFORMATION: /gene= "AR"
OTHER INFORMATION: /number= 14
FEATURE:
NAME/KEY: exon
LOCATION: complement (29945..30073)
OTHER INFORMATION: /gene= "AR"
OTHER INFORMATION: /number= 15
FEATURE:
NAME/KEY: exon
LOCATION: complement (29664..29856)
OTHER INFORMATION: /gene= "ARS"
OTHER INFORMATION: /number= 16
```

US-08-814-095-7

```
Query Match      19.7%;  Score 197.2;  DB 3;  Length 35060;
Best Local Similarity    61.9%;  Pred. No. 9.9e-43;
Matches  370;  Conservative  0;  Mismatches 213;  Indels  15;  Gaps  3;

Qy       1 CCTGCCACCATGCCTGGCTAATTTTCTTTATTTTTAGTAGAGACGGAGGTTTTGCCATGTTG    60
           || |||| ||||| |||||           ||||||||||| ||| |||||||||||||||
Db    9942 CCCACCACAATGCCCAGCTAATTTTTTATATAATTTCAGTAGAGACAGGGTTTCACCATGTTG 10001

Qy      61 ACCAGGCTGGTCTCGAACTCTTGACCTCAGGTGATCCGCCCGCCTCAGCGCCTCCCAAAGTG   120
           |||||||||||||||||| || || ||||| ||||| |||||| |||||||||||||||||
Db   10002 GCCAGGCTGGTCTTCTCAAAATCCTGGCCTCAAGTGATCTGCCCACCTCAGCCTCCCAAAGTG 10061
```

```
Qy   121 CTGGGATTATATAGGCGGTGAGCCGCCGCGACCAGCCAACACATTTTTAAATACTGAAAGTAG  180
         |||||||||  ||  ||||||||  |||  ||| ||||  || ||      ||||| |||||||
Db 10062 CTGGGATTACAGATGTGAGCCACCTCACTCAGCCAGAATCCCTTCCAAATTAAAAATCAA   10121

Qy   181 AGGGAATAGTTATATAGTGTACCCCAT------TTACCCATCACTCAGTTTCAACAGCTGG   233
         | | |||  ||  || |||  |||  ||      |  || ||| ||| |||  | |||||||
Db 10122 ACAGAAACCCACAGAATACACATTGTTTGGCAATTGCTCTAGTAATATGTGCATCTTGGG   10181

Qy   234 TGACATATATTATTTCTTCTATACCAGTACCGTACTCTCCCCCACTGGGATTATTTTAAGGC   293
         || |   | |  |||| |  | || || || | |  |   ||||| |||||  || ||| ||
Db 10182 TGCAAGCGTCTAAATTGCTAAACCTGCATCTGCAGTTTTTACATTTTAACATGCTTCCGGC   10241

Qy   294 AAAACCCAGA--TGACATTTATATCCCTAAATACTTTAGATAAAAGGTGTTCTTTGAAAAA   350
         ||  |||  |  || |  ||| || |||  |||     || ||||| ||| | ||| ||||
Db 10242 CAGGCGTGGTGGGCTCACGCCTGTAATCCCAGCACTTTGGGAAGCCGAGGCAGGTGGATCA   10301

Qy   351 AATCATAAC------CTCAGACCACAGCCTGCCAACATGGTGAAACCCTGTCTGTTACTAAA   405
         || ||| ||      |||||   | ||||| |||||| | |||||||| |  | ||||||||
Db 10302 CTTCAGGTCAGGAGTTCGAGACTAGCCTGACCAACATAGTGAAACCCCATCCTCTACTAAA   10361

Qy   406 AATACAAAAATTAGCTTGGCATCGTCGTGGGCACCTGTAATCCCAGCTACTCACTCAGGAAGCT   465
         || | |||||||| |  ||| | | | ||| ||| ||||||||||||||||  | |||||||||
Db 10362 AAATACAAAAATTACCCAGTGTGGTGGCGCCATGCCTGTAATCCCAGCTACTCAGGAGGCT   10421

Qy   466 GAGGCAGGAGAATCACTTGAAATCCGGGAAGCAGAGAATTGCAGTGAGCCGAGATTGCAGTC   525
         ||||||||||||| |||||| | ||||||| |||||| || ||||||||||||||||| | |
Db 10422 GAGGCAGGAGAATTGCTTGAACCCGGGAGGCAGAGATTGTAGTGAGCCGAGATTGCACCA   10481

Qy   526 GAGCCTGGGCGACAGAGACGAGAAATGAAAACTCTGTCTCAAAAAACAAAACAAAACAAAAA   583
         ||||| |||| |||| | | || |  |||||||| |||||||| || |||| | || ||| |
Db 10482 TGGCACTCCATCCTGGGCAACAGGAGCAAAAACTCCGTCTCAAACAAACAAAAACAAACA   10539
```

RESULT 6
US-09-345-882-1
Sequence 1, Application US/09345882
Patent No. 6399373
GENERAL INFORMATION:
APPLICANT: Bougueleret, Lydie
TITLE OF INVENTION: A NUCLEIC ACID ENCODING A RETINOBLASTOMA BINDING PROTEIN (RBP-7)
TITLE OF INVENTION: AND POLYMORPHIC MARKERS ASSOCIATED WITH SAID NUCLEIC ACID.
FILE REFERENCE: GENSET.031A
CURRENT APPLICATION NUMBER: US/09/345,882
CURRENT FILING DATE: 1999-06-30
PRIOR APPLICATION NUMBER: US 60/091,315
PRIOR FILING DATE: 1998-06-30
PRIOR APPLICATION NUMBER: US 60/111,909
PRIOR FILING DATE: 1998-12-10
NUMBER OF SEQ ID NOS: 140
SOFTWARE: Patent.pm
SEQ ID NO 1
LENGTH: 162450
TYPE: DNA
ORGANISM: Homo sapiens
FEATURE:
NAME/KEY: allele
LOCATION: 72794
OTHER INFORMATION: 5-124-273 : polymorphic base A or G
FEATURE:
NAME/KEY: allele
LOCATION: 88073
OTHER INFORMATION: 5-127-261 : polymorphic base A or C
FEATURE:
NAME/KEY: allele
LOCATION: 90842
OTHER INFORMATION: 99-1437-325 : polymorphic base A or G
FEATURE:
NAME/KEY: allele
LOCATION: 93714
OTHER INFORMATION: 5-128-60 : polymorphic base deletion of GT
FEATURE:
NAME/KEY: allele

FEATURE:
NAME/KEY: allele
LOCATION: 72771..72817
OTHER INFORMATION: polymorphic fragment 5-124-273 SEQ ID30
FEATURE:
NAME/KEY: allele
LOCATION: 72771..72817

OTHER INFORMATION: 97122
OTHER INFORMATION: 99-1442-224 : polymorphic base G or T
FEATURE:
NAME/KEY: allele
LOCATION: 97152
OTHER INFORMATION: 5-129-144 : polymorphic base deletion of T
FEATURE:
NAME/KEY: allele
LOCATION: 99098
OTHER INFORMATION: 5-130-257 : polymorphic base A or G
FEATURE:
NAME/KEY: allele
LOCATION: 99117
OTHER INFORMATION: 5-130-276 : polymorphic base A or G
FEATURE:
NAME/KEY: allele
LOCATION: 103806
OTHER INFORMATION: 5-131-395 : polymorphic base A or T
FEATURE:
NAME/KEY: allele
LOCATION: 106940
OTHER INFORMATION: 5-133-375 : polymorphic base insertion of A
FEATURE:
NAME/KEY: allele
LOCATION: 108106
OTHER INFORMATION: 5-135-155 : polymorphic base insertion of A
FEATURE:
NAME/KEY: allele
LOCATION: 108149
OTHER INFORMATION: 5-135-198 : polymorphic base insertion of GTTT
FEATURE:
NAME/KEY: allele
LOCATION: 108308
OTHER INFORMATION: 5-135-357 : polymorphic base A or G
FEATURE:
NAME/KEY: allele
LOCATION: 108471
OTHER INFORMATION: 5-136-174 : polymorphic base C or T
FEATURE:
NAME/KEY: allele
LOCATION: 134134
OTHER INFORMATION: 5-140-120 : polymorphic base C or T
FEATURE:
NAME/KEY: allele
LOCATION: 134362
OTHER INFORMATION: 5-140-348 : polymorphic base insertion of A
FEATURE:
NAME/KEY: allele
LOCATION: 134374
OTHER INFORMATION: 5-140-361 : polymorphic base insertion of CA
FEATURE:
NAME/KEY: allele
LOCATION: 146328
OTHER INFORMATION: 5-143-84 : polymorphic base A or G
FEATURE:
NAME/KEY: allele
LOCATION: 146345
OTHER INFORMATION: 5-143-101 : polymorphic base A or C
FEATURE:
NAME/KEY: allele
LOCATION: 150329
OTHER INFORMATION: 5-145-24 : polymorphic base A or G
FEATURE:
NAME/KEY: allele
LOCATION: 160031
OTHER INFORMATION: 5-148-352 : polymorphic base G or T
FEATURE:
NAME/KEY: allele
LOCATION: 72771..72817

```
FEATURE:
NAME/KEY: allele
LOCATION: 108084..108130
OTHER INFORMATION: polymorphic fragment 5-135-155 SEQ ID38
FEATURE:
NAME/KEY: allele
LOCATION: 108084..108130
OTHER INFORMATION: polymorphic fragment 5-135-155 SEQ ID59
FEATURE:
NAME/KEY: allele
LOCATION: 108127..108177
OTHER INFORMATION: polymorphic fragment 5-135-198 SEQ ID39
FEATURE:
NAME/KEY: allele
LOCATION: 108127..108177
OTHER INFORMATION: polymorphic fragment 5-135-198 SEQ ID60
FEATURE:
```

```
Query Match                 19.0%;    Score 189.8; DB 4;  Length 162450;
Best Local Similarity       60.3%;    Pred. No. 1.6e-40;
Matches   359; Conservative     0; Mismatches  217; Indels   19; Gaps    2;

Qy          1 CCTGCCACCCATGCCTGGCTGTAATTTCTTATTTTAGTAGAGACGAGGTTTTTGCATGTTG      60
              || ||||||||||||||||||||||||| ||| |||||||||||||||||||||||||| |||
Db      86234 CCCGCCACCCACTGCCTGGCTGTAATTTTTGTATTTTAGTAGAGACGAGGTTTCACCATTTG   86293

Qy         61 ACCAGGCGTGGTCTCGAACTCTTGACCTCAGGTGATCCGCCTGCCTCAGCGCTCCCAAAGTG     120
              |||||||| ||||||||||| || |||| ||||||||||||||||||||||| |||||||||
Db      86294 GCCAGGCCTGGTCTCGAACTCCTCAGGTGATCCAGCCTGCCTCCGCTTCCGCTCCCAAAGTG   86353

Qy        121 CTGGGATTATAGGCGTGAGCCGCCGCACCCAGCCAACACATTTTTAAATACTGAAAAGTAG     180
              |||||| | |||||||||||||||  ||||||||||||||||||||||||||||||||||
Db      86354 CTGGGAGTACAATACTTCCAGCCAAGTGACACTTCTTATTCAAATAGCAAGCTTTGGAAG   86413

Qy        181 AGGGAAT--------------AGTTTATAGTGTGTACCCCATTTACCCATTCACTCAGTTTCA     225
                  |  |          ||||||||||| ||| |||||||||||||||
Db      86414 ATTTTGCTTTTGTTTCTGCCAGGGAACAAATTTTTAAGTTATTTGACTGACATTGAAAATCAA   86473

Qy        226 ACAGCTGGTGACATATTTTATTTCTTCTATACCAGTACCGTACTCTCCCCCACTGGGGATTAT     285
              | |||| |||||||||||||||||||||||||||||||||||||||||||||||| |||| ||
Db      86474 GGGGGTGGTTGATTCCCCCATGTCACTTTAAAAAGCTTAATATTATTAATAGTTTTTTACTT   86533

Qy        286 TTTAAGGCAAAAACCCAGATGACA------TTTTATCCCTAAAATACTTTAGATAAAGGTGTTC     341
              |||| || || ||                 |||| |||||||||||||||||||||||||
Db      86534 TTTTGGCCCAGGTGGGGTGGCTCACGCCTTATAAATCCCAGCACTTTGGGAGGCGGAGGTGGGC   86593

Qy        342 TTTGAAAAAATCATAACCTCAGACACCAGCCTGCCAACATGTGAAACCCCGTCTGTAC     401
              ||| ||||||||||||||||||||||||||||||||||||||||||||||| ||||||
Db      86594 GGATCACGAGGTCAGGAGATCGAGACTATCCTGGCTAACATGGTGAAACCCCTTCTCTAC   86653

Qy        402 TAAAAATACAAAAAAATTAGCTTGGCATGGTGGCGGGCGCCTGTAGTCCCAGCTACTCAGGA     461
                |  | ||||||||| ||||||| ||||||||| ||||||||||||||||||| |||
Db      86654 CAAAAATACAAAAATTAGCTGGGCGGTGGTGGCGCGCACCTGTAATCCCAGCTACTCAGGA   86713

Qy        462 AGCTGAGGCAGGAGAATCACTTGAATCCGGGAAGCACAGAGATTGCAGTGAGCTGAGATTGC     521
              ||| ||||||||||||||||| ||||||||||||||||||| ||||| |||||| |||| ||
Db      86714 GGCTGAGGCAGGAGAATTGCTTGAACCCAGGAGGCGGAGGTTGCAGTGAGCCAAGGTGGT   86773

Qy        522 AGTCGAGCCTGGGCGACAGAGACAGAAATGAAAACTCTGTCTCAAAAACAAACAA     576
              ||    |||||||||||||||||||||||||||||||||| |||||||||||||
Db      86774 GCCACTGCACTCCAGCCCGGGCAACAGAGCGAGACTCCGCCTCAAAAAAGAAAAAA   86828
```

RESULT 7
US-08-814-095-7/c
Sequence 7, Application US/08814095
Patent No. 6025183

```
FEATURE:
NAME/KEY: allele
LOCATION: 88050..88096
OTHER INFORMATION: polymorphic fragment 5-124-273 SEQ ID51
FEATURE:
NAME/KEY: allele
LOCATION: 88050..88096
OTHER INFORMATION: polymorphic fragment 5-127-261 SEQ ID31
FEATURE:
NAME/KEY: allele
LOCATION: 90819..90865
OTHER INFORMATION: polymorphic fragment 5-127-261 SEQ ID52
FEATURE:
NAME/KEY: allele
LOCATION: 90819..90865
OTHER INFORMATION: complement polymorphic fragment 99-1437-325 SEQ ID49
FEATURE:
NAME/KEY: allele
LOCATION: 90819..90865
OTHER INFORMATION: complement polymorphic fragment 99-1437-325 SEQ ID70
FEATURE:
NAME/KEY: allele
LOCATION: 93690..93736
OTHER INFORMATION: polymorphic fragment 5-128-60 SEQ ID53
FEATURE:
NAME/KEY: allele
LOCATION: 93690..93736
OTHER INFORMATION: polymorphic fragment 5-128-60 SEQ ID32
FEATURE:
NAME/KEY: allele
LOCATION: 97099..97145
OTHER INFORMATION: polymorphic fragment 99-1442-224 SEQ ID50
FEATURE:
NAME/KEY: allele
LOCATION: 97099..97145
OTHER INFORMATION: polymorphic fragment 99-1442-224 SEQ ID71
FEATURE:
NAME/KEY: allele
LOCATION: 97130..97177
OTHER INFORMATION: polymorphic fragment 5-129-144 SEQ ID33
FEATURE:
NAME/KEY: allele
LOCATION: 97130..97177
OTHER INFORMATION: polymorphic fragment 5-129-144 SEQ ID54
FEATURE:
NAME/KEY: allele
LOCATION: 99075..99121
OTHER INFORMATION: polymorphic fragment 5-130-257 SEQ ID34
FEATURE:
NAME/KEY: allele
LOCATION: 99075..99121
OTHER INFORMATION: polymorphic fragment 5-130-257 SEQ ID55
FEATURE:
NAME/KEY: allele
LOCATION: 99094..99140
OTHER INFORMATION: polymorphic fragment 5-130-276 SEQ ID56
FEATURE:
NAME/KEY: allele
LOCATION: 99094..99140
OTHER INFORMATION: polymorphic fragment 5-130-276 SEQ ID35
FEATURE:
NAME/KEY: allele
LOCATION: 103783..103828
OTHER INFORMATION: polymorphic fragment 5-131-395 SEQ ID36
FEATURE:
NAME/KEY: allele
LOCATION: 103783..103828
OTHER INFORMATION: polymorphic fragment 5-131-395 SEQ ID57
FEATURE:
NAME/KEY: allele
LOCATION: 106918..106966
OTHER INFORMATION: polymorphic fragment 5-133-375 SEQ ID37
FEATURE:
NAME/KEY: allele
LOCATION: 106918..106966
OTHER INFORMATION: polymorphic fragment 5-133-375 SEQ ID58
```

NUMBER OF SEQUENCES: 7
CORRESPONDENCE ADDRESS:
ADDRESSEE: KOHN & ASSOCIATES
STREET: 30500 No. 602518 3thwestern Highway, Suite 410
CITY: Farmington Hills
STATE: Michigan
COUNTRY: U.S.
ZIP: 48334
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/814,095
FILING DATE:
CLASSIFICATION: 800
ATTORNEY/AGENT INFORMATION:
NAME: Montgomery, Ilene N.
REGISTRATION NUMBER: 38,972
REFERENCE/DOCKET NUMBER: 2391.00066
TELECOMMUNICATION INFORMATION:
TELEPHONE: (248) 539-5050
TELEFAX: (248) 539-5055
INFORMATION FOR SEQ ID NO: 7:
SEQUENCE CHARACTERISTICS:
LENGTH: 35060 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: other nucleic acid
DESCRIPTION: /desc = "Cosmid including ACHE
DESCRIPTION: promotor, ACHE gene and ARS gene"
HYPOTHETICAL: NO
ANTI-SENSE: NO
ORIGINAL SOURCE:
ORGANISM: Homo sapiens
POSITION IN GENOME:
CHROMOSOME/SEGMENT: 7q22
FEATURE:
NAME/KEY: promoter
LOCATION: 4089..22464
OTHER INFORMATION: /function= "ACHE Promotor"
OTHER INFORMATION: /standard_name= "ACHE Promotor"
FEATURE:
NAME/KEY: exon
LOCATION: 22465..22537
OTHER INFORMATION: /function= "non-translated"
OTHER INFORMATION: /gene= "ACHE"
OTHER INFORMATION: /number= 1
FEATURE:
NAME/KEY: exon
LOCATION: 24090..25177
IDENTIFICATION METHOD: experimental
OTHER INFORMATION: /function= "(translation start:
OTHER INFORMATION: 24110)"
OTHER INFORMATION: /gene= "ACHE"
OTHER INFORMATION: /number= 2
FEATURE:
NAME/KEY: exon
LOCATION: 25524..26009
IDENTIFICATION METHOD: experimental
OTHER INFORMATION: /evidence= EXPERIMENTAL
OTHER INFORMATION: /gene= "ACHE"
OTHER INFORMATION: /number= 3
FEATURE:
NAME/KEY: exon
LOCATION: 27005..27274
IDENTIFICATION METHOD: experimental
OTHER INFORMATION: /evidence= EXPERIMENTAL
OTHER INFORMATION: /gene= "ACHE"
OTHER INFORMATION: /number= 4

FEATURE:
NAME/KEY: exon
LOCATION: 27255..28007
IDENTIFICATION METHOD: experimental
OTHER INFORMATION: /evidence= EXPERIMENTAL
OTHER INFORMATION: /gene= "ACHE"
OTHER INFORMATION: /number= 5
FEATURE:
NAME/KEY: exon
LOCATION: 27385..27387
FEATURE:
NAME/KEY: exon
LOCATION: 28008..28129
IDENTIFICATION METHOD: experimental
OTHER INFORMATION: /evidence= EXPERIMENTAL
OTHER INFORMATION: /gene= "ACHE"
OTHER INFORMATION: /number= 6
FEATURE:
NAME/KEY: terminator
LOCATION: 28129..28131
FEATURE:
NAME/KEY: exon
LOCATION: complement (34528..34895)
OTHER INFORMATION: /function= "arsenite resistance
OTHER INFORMATION: gene"
OTHER INFORMATION: /gene= "AR"
OTHER INFORMATION: /number= 1
FEATURE:
NAME/KEY: exon
LOCATION: complement (34092..34358)
OTHER INFORMATION: /gene= "AR"
OTHER INFORMATION: /number= 2
FEATURE:
NAME/KEY: exon
LOCATION: complement (33779..33963)
OTHER INFORMATION: /gene= "AR"
OTHER INFORMATION: /number= 3
FEATURE:
NAME/KEY: exon
LOCATION: complement (33493..33591)
OTHER INFORMATION: /gene= "AR"
OTHER INFORMATION: /number= 4
FEATURE:
NAME/KEY: exon
LOCATION: complement (33297..33408)
OTHER INFORMATION: /gene= "AR"
OTHER INFORMATION: /number= 5
FEATURE:
NAME/KEY: exon
LOCATION: complement (32959..33094)
OTHER INFORMATION: /gene= "AR"
OTHER INFORMATION: /number= 6
FEATURE:
NAME/KEY: exon
LOCATION: complement (32386..32468)
OTHER INFORMATION: /gene= "AR"
OTHER INFORMATION: /number= 7
FEATURE:
NAME/KEY: exon
LOCATION: complement (32569..32628)
OTHER INFORMATION: /gene= "AR"
OTHER INFORMATION: /number= 8
FEATURE:
NAME/KEY: exon
LOCATION: complement (31894..32080)
OTHER INFORMATION: /gene= "AR"
OTHER INFORMATION: /number= 9
FEATURE:
NAME/KEY: exon
LOCATION: complement (31363..31534)
OTHER INFORMATION: /gene= "AR"
OTHER INFORMATION: /number= 10
FEATURE:

```
;   NAME/KEY:           exon
;   LOCATION:           complement (31131..31284)
;   OTHER INFORMATION:  /gene= "AR"
;   OTHER INFORMATION:  /number= 11
;   FEATURE:
;   NAME/KEY:           exon
;   LOCATION:           complement (30816..31011)
;   OTHER INFORMATION:  /gene= "AR"
;   OTHER INFORMATION:  /number= 12
;   FEATURE:
;   NAME/KEY:           exon
;   LOCATION:           complement (30470..30626)
;   OTHER INFORMATION:  /gene= "AR"
;   OTHER INFORMATION:  /number= 13
;   FEATURE:
;   NAME/KEY:           exon
;   LOCATION:           complement (30187..30274)
;   OTHER INFORMATION:  /gene= "AR"
;   OTHER INFORMATION:  /number= 14
;   FEATURE:
;   NAME/KEY:           exon
;   LOCATION:           complement (29945..30073)
;   OTHER INFORMATION:  /gene= "AR"
;   OTHER INFORMATION:  /number= 15
;   FEATURE:
;   NAME/KEY:           exon
;   LOCATION:           complement (29664..29856)
;   OTHER INFORMATION:  /gene= "ARS"
;   OTHER INFORMATION:  /number= 16
;   US-08-814-095-7
```

```
Query Match         18.9%;  Score 189;  DB 3;  Length 35060;
Best Local Similarity    62.1%;
Matches  373;  Conservative   0;  Mismatches  210;  Indels   18;  Gaps   4;
```

```
Qy      1 CCTGCCACCATGCCTCGGCTAATTTCTTATTTTTAGTAGAGACGAGGTTTTGCCATGTTG    60
          ||||||||||||||||||| ||||||||||||  ||||||||||||||||||||||||||||
Db  19913 CCTGCCACCACACCTGGCTAATTTTTGTATATTTTAGCAGAGGTGGGGTTTCACCATGTTG  19854

Qy     61 ACCAGGCTGGTCTCGAACTCTTGACCTCAGGTGATCCGCCTGCCTCAGCCTCCCAAAGTG   120
          ||||||||| ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  19853 GCCAGGCTGGTCTCGAACTCCTGACCTCAGGTGATCCTCCTGCCTCAGCCTCCCAAAGTG  19794

Qy    121 CTGGGATTATAGGCGTGCAGCGCGCGACC------CAGCCAACACATTTTTAAATACTG   172
          |||||||||  |||  ||||| || |||||       |||||||||||  ||||||||||
Db  19793 TTGGGATTACAGGCGTGAGCTGCCCATGCCTGGCAGGGAAACCCCACTTCTATTACAGACCT  19734

Qy    173 AAAAGTAGAGGAGGAATAGTTATAGTGTACCCCATTTACCCATCACTCAGTTTCAACAGCTG   232
          ||   |  || || || |||||||||||||||||| |  | | | |||||||||||||||
Db  19733 TGATCCACATCTTCTAACACAAAAGATTCTTTATTTCCTCTATTTATTGGTTAATATTTCAA  19674

Qy    233 GTGACATATTTATTTCTTCTATACCAGTACCGTACTCTCCCCACTGGGGATTATTTTAAGG   292
          | | ||                   |||||||  |   |   || | |    || ||||||
Db  19673 TTCAGCTGTATGAGAGACCAAAAAACCAAAATAATAGGCTTTAACAGAGATATAATAGAGGCCGGG  19614

Qy    293 CAAAACCCAGATGACATTTTATCCCTAAATACTTTAGATAAAGGTGTTCTTTTTGAAA-----   348
          ||     |   |  |  | | | |  | | |   |||||||||||||||||||||||
Db  19613 CATGGTGGCTCATGCCTGTTTGTTGTAATCCCAGTATTTTTGGGAGGCTGAGGCAGGCAGATCAC  19554

Qy    349 --AAAATCATAAACCTCAGGAGCCACAGGCCCAACAATGGTGAAAACCCTGTTCCTGTTACTAAAA   406
          || || |     | ||| |   |  |||  ||  | ||  | |   ||  | ||| ||| |||||
Db  19553 CTGAGGTTGGGAGTTCAAGACCAGCCTGACCAACATGGAGAAAACCCTGTCTCTACTAAAA  19494

Qy    407 ATAC--AAAAATTAGCTTGGGCATGGTGGTCGTGGGCCACCTGTAATCCCAGCTACTCAGGAAGC   464
          ||||    ||||||||||||||||||||||||||||||  |||||||||||||||||||||||
Db  19493 ATACAAAAAAAATTAGCTGGGCATGGTGGCGGGCGCCTGTAATCCCAGCTACTCAGGAGGC  19434

Qy    465 TGAGGCAGGAGAATCACTTGAATCCCGGGAAGCAGAGAGTTGCAGTGAGCTGAGAT--TGCA   522
          |||||||||||||||| |||||| | ||| ||||||||||||||||||||||||      ||
Db  19433 TGAGGCAGGAGAATCGCTTGAACCCAGGAGGTGGAGGGTGCAGTGAGCCGAGATCATGCC  19374

Qy    523 GTCGAGCCTGGGCGACAGAGCGAGACAGAGAAATGAAACTCTGTCTCAAAAACAAACAAAAA   582
          || || |  ||  ||| | || |  ||| | | |||| ||  | ||| ||||  |||||
Db  19373 ATTGTACTCCAGCCTGGGCAACAAGAGCGAGGGAAACTCTGTCTCAAAAAAAAAAAAAATAAT  19314
```

```
Db  19373 ATTGTACTCCAGCCTGGGCAACAAGAGCGAGGGAAACTCTGTCTCAAAAAAAAAAAAAATAAT  19314
Qy    583 A  583
          |
Db  19313 A  19313
```

```
;       NAME/KEY: exon
;       LOCATION:
US-07-906-871-15            16397..17327

Matches 339;  Conservative   0;  Mismatches 231;  Indels    4;  Gaps    1;
Best Local Similarity   59.1%;  Pred. No. 5.7e-40;
Query Match   18.6%;  Score 186.4;  DB 1;  Length 17327;

Qy    22 TTTTCTTATTTTTAGTAGAGACGAGAGGTTTTGCCCATGTTGGACCAGGCTGGTCTCGAACTCT  81
         |||||| |||| ||||||||||||||| ||||||||||||||||||||||||||||||||||||
Db  7274 TTTTCTCTTTTTTGGTTAGAAAATGGGAGTCTCTCTATGTTGCCCAGGCTGGTCTCGAAACTCC 7333

Qy    82 TGACCTCAGGTGATCCGGCCGCCTCAGCCTCCCAAAGTGCTGGGATTATAGGGGTGAGCC  141
         || ||||||||| ||| |||| ||| ||||||||||| || ||||||||| |||| ||||
Db  7334 TGGGCTCAAGTGATCCTCCCCACCTCAGTCTCCCAAAGTGCTGGGATTTTCAGGCATGAGCT 7393

Qy   142 GCCGCACCCAGCCAACACATTTTTTAAATACTGAAAAGTAGAGGGAATAGTTATAGTGTACC  201
         |||  |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  7394 ACCACACCCAACCTAATTTTTTTATTTTTTATTAAAAAAAGTTTTTGCCTACCCG 7453

Qy   202 CCATTTACCCATCACTCAGTTCAACAGCTGGTGACATATATTTTATTTTTTTCTTATACCAGTA  261
         || |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  7454 CCTCTCACCCTCTCACTCATTTTTTTAAGTTATAGGTTTTTCTACAATGCCACATTGCTCTTTT 7513

Qy   262 CCGTACTTCTCCCACTGGGGATTATTTTAAGGCAAAAACCCAGATGACATTTTATCCCTAAA  321
         || ||| ||||||||||||||||||| |||||||||||||||||||||||||||||||
Db  7514 CCATAAAAAGTACTTTCCGCCGGGTGTGGTGGCTCACACCTGTAATCCCAGCACTTTGGG 7573

Qy   322 TACTTTAGATAAAGGTGTTCTTTTGAAAAAAAAATCATAACCTCAGGACCAGCCTGGCCAACA  381
         || ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  7574 AGACTGAGG----CGGGTGGATCACCTGAGGTCGAGGAGTTCGAGACCAGCCTGGCCAACA 7629

Qy   382 TGGTGAAACCCCTGCTCTGTCTACTAAAAAATACAAAAAATTAGCTTGGCATGGTGGTCGTGGGCACCT  441
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  7630 TGGTGAAACCCCATCCCTACTAAAAATACAAAAATTAGCCTGGGCATGGTGGTGGGCGCCT 7689

Qy   442 GTAATCCCAGCTACTCAGGAAGCTGAGGCAGGAGAATCACTTGAATCCCGGGAAGCAGAGA  501
         |||||||||||||||||||||||||||||||||||||| ||||||||||||
Db  7690 GTCATCCCAGCTACTCGGGAGGCTGAGGCAGGAGAATCGCTTGAACCTGGGAGGCGGAGG 7749

Qy   502 TTGCAGTGAGCTGAGATTGCAGTCGAGCCTGGGCGACAGAGCAAGAATGAAACTCTGTC  561
         ||||||||||||| || || || ||||||||||||
Db  7750 TTGCAGTGAGCCGAGATTGTGCCACTGCAGTCCAGCCTGGGCAACAGAGCGAGACTTCAT 7809

Qy   562 TCAAAAACAAAACAAAACAAAAAAAAACCACTATACAT  595
         |||  ||| |||| |||||||
Db  7810 CTCAAAAAAAAAAAAAAAGTACTTTTCTTCAT 7843
```

RESULT 9
US-09-345-882-1/c
Sequence 1, Application US/09345882
Patent No. 6399373

```
;       OTHER INFORMATION:
;       FEATURE:
;       NAME/KEY: allele
;       LOCATION: 88073
;       OTHER INFORMATION: 93714           :  polymorphic base A or G
;       FEATURE:
;       NAME/KEY: allele
;       LOCATION: 90842
;       OTHER INFORMATION: 99-1437-325      :  polymorphic base A or C
;       FEATURE:
;       NAME/KEY: allele
;       LOCATION: 97122
;       OTHER INFORMATION: 99-1442-224      :  polymorphic base G or T
;       FEATURE:
;       NAME/KEY: allele
;       LOCATION: 97152
;       OTHER INFORMATION: 5-128-60         :  polymorphic base deletion of GT
;       FEATURE:
;       NAME/KEY: allele
;       LOCATION: 99098
;       OTHER INFORMATION: 5-129-144        :  polymorphic base deletion of T
;       FEATURE:
;       NAME/KEY: allele
;       LOCATION: 99117
;       OTHER INFORMATION: 5-130-257        :  polymorphic base A or G
;       FEATURE:
;       NAME/KEY: allele
;       LOCATION: 103806
;       OTHER INFORMATION: 5-130-276        :  polymorphic base A or G
;       FEATURE:
;       NAME/KEY: allele
;       LOCATION: 108149
;       OTHER INFORMATION: 5-131-395        :  polymorphic base A or T
;       FEATURE:
;       NAME/KEY: allele
;       LOCATION: 106940
;       OTHER INFORMATION: 5-133-375        :  polymorphic base insertion of A
;       FEATURE:
;       NAME/KEY: allele
;       LOCATION: 108106
;       OTHER INFORMATION: 5-135-155        :  polymorphic base insertion of A
;       FEATURE:
;       NAME/KEY: allele
;       LOCATION: 108308
;       OTHER INFORMATION: 5-135-198        :  polymorphic base insertion of GTTT
;       FEATURE:
;       NAME/KEY: allele
;       LOCATION: 134362
;       OTHER INFORMATION: 5-135-357        :  polymorphic base A or G
;       FEATURE:
;       NAME/KEY: allele
;       LOCATION: 108471
;       OTHER INFORMATION: 5-136-174        :  polymorphic base C or T
;       FEATURE:
;       NAME/KEY: allele
;       LOCATION: 134134
;       OTHER INFORMATION: 5-140-120        :  polymorphic base C or T
;       FEATURE:
;       NAME/KEY: allele
;       LOCATION: 134374
;       OTHER INFORMATION: 5-140-361        :  polymorphic base insertion of CA
;       FEATURE:
;       NAME/KEY: allele
;       LOCATION: 146328
;       OTHER INFORMATION: 5-140-348        :  polymorphic base insertion of A
;       FEATURE:
;       NAME/KEY: allele
;       LOCATION: 146345
;       OTHER INFORMATION: 5-143-84         :  polymorphic base A or G
;       FEATURE:
;       NAME/KEY: allele
;       LOCATION: 146345
;       OTHER INFORMATION: 5-143-101        :  polymorphic base A or C
```

```
FEATURE:
NAME/KEY: allele
LOCATION: 150329
OTHER INFORMATION: 5-145-24    :  polymorphic base A or G
FEATURE:
NAME/KEY: allele
LOCATION: 160031
OTHER INFORMATION: 5-148-352   :  polymorphic base G or T
FEATURE:
NAME/KEY: allele
LOCATION: 72771..72817
OTHER INFORMATION: polymorphic fragment 5-124-273 SEQ ID30
FEATURE:
NAME/KEY: allele
LOCATION: 72771..72817
OTHER INFORMATION: polymorphic fragment 5-124-273 SEQ ID51
FEATURE:
NAME/KEY: allele
LOCATION: 88050..88096
OTHER INFORMATION: polymorphic fragment 5-127-261 SEQ ID31
FEATURE:
NAME/KEY: allele
LOCATION: 88050..88096
OTHER INFORMATION: polymorphic fragment 5-127-261 SEQ ID52
FEATURE:
NAME/KEY: allele
LOCATION: 90819..90865
OTHER INFORMATION: complement polymorphic fragment 99-1437-325 SEQ ID49
FEATURE:
NAME/KEY: allele
LOCATION: 90819..90865
OTHER INFORMATION: complement polymorphic fragment 99-1437-325 SEQ ID70
FEATURE:
NAME/KEY: allele
LOCATION: 93690..93736
OTHER INFORMATION: polymorphic fragment 5-128-60 SEQ ID32
FEATURE:
NAME/KEY: allele
LOCATION: 93690..93736
OTHER INFORMATION: polymorphic fragment 5-128-60 SEQ ID53
FEATURE:
NAME/KEY: allele
LOCATION: 97099..97145
OTHER INFORMATION: polymorphic fragment 99-1442-224 SEQ ID71
FEATURE:
NAME/KEY: allele
LOCATION: 97099..97145
OTHER INFORMATION: polymorphic fragment 99-1442-224 SEQ ID50
FEATURE:
NAME/KEY: allele
LOCATION: 97130..97177
OTHER INFORMATION: polymorphic fragment 99-1442-224 SEQ ID33
FEATURE:
NAME/KEY: allele
LOCATION: 97130..97177
OTHER INFORMATION: polymorphic fragment 5-129-144 SEQ ID54
FEATURE:
NAME/KEY: allele
LOCATION: 99075..99121
OTHER INFORMATION: polymorphic fragment 5-129-144 SEQ ID34
FEATURE:
NAME/KEY: allele
LOCATION: 99075..99121
OTHER INFORMATION: polymorphic fragment 5-130-257 SEQ ID55
FEATURE:
NAME/KEY: allele
LOCATION: 99094..99140
OTHER INFORMATION: polymorphic fragment 5-130-257 SEQ ID35
FEATURE:
NAME/KEY: allele
LOCATION: 99094..99140
OTHER INFORMATION: polymorphic fragment 5-130-276 SEQ ID56
FEATURE:
NAME/KEY: allele
LOCATION: 103783..103828
OTHER INFORMATION: polymorphic fragment 5-130-276 SEQ ID36
FEATURE:
NAME/KEY: allele
LOCATION: 103783..103828
OTHER INFORMATION: polymorphic fragment 5-131-395 SEQ ID57
FEATURE:
NAME/KEY: allele
LOCATION: 106918..106966
OTHER INFORMATION: polymorphic fragment 5-131-395 SEQ ID37
FEATURE:
NAME/KEY: allele
LOCATION: 106918..106966
OTHER INFORMATION: polymorphic fragment 5-133-375 SEQ ID58
FEATURE:
NAME/KEY: allele
LOCATION: 108084..108130
OTHER INFORMATION: polymorphic fragment 5-133-375 SEQ ID38
FEATURE:
NAME/KEY: allele
LOCATION: 108084..108130
OTHER INFORMATION: polymorphic fragment 5-135-155 SEQ ID59
FEATURE:
NAME/KEY: allele
LOCATION: 108127..108177
OTHER INFORMATION: polymorphic fragment 5-135-155 SEQ ID39
FEATURE:
NAME/KEY: allele
LOCATION: 108127..108177
OTHER INFORMATION: polymorphic fragment 5-135-198 SEQ ID60
FEATURE:
```

```
Query Match      18.1%;  Score 181;  DB 4;  Length 162450;
Best Local Similarity   62.5%;  Pred. No. 3.6e-38;
Matches   377;  Conservative    0;  Mismatches 200;  Indels   26;  Gaps   5;

Qy      1  CCTGCCACCATGGCCTGGGCTAATTTCTTATTTTTTAGTAGAGACGAGGTTTGCCATGTTG  60
           ||||||||| | |||||||||||||||||||     ||||||||||||||||| ||||||||
Db  86688  CCTGCCACCGCCCCAGCCTAATTTTGTTATTTTTTGGTTTGAGAAGACGGGTTTCACCATGTTA  86629

Qy     61  ACCAGGGCTGGTCTCGAACTCTTGACCTCAGGTGATCCGCTGCCTCAGCCTCCCAAAGTG  120
           ||||| | |||||||||||||||| | || || |||||||||||||||||||||||||||
Db  86628  GCCAGGGATAGTCTCGATCTCCTGACCTC--GTGATCGCCCACCTC-GCCTCCCAAAGTG  86572

Qy    121  CTGGGATTATAGGCGTGAGCCGCGCGCCACCCAGCCAACATTTTTTAAATACTGAAAAGTAG  180
           |||||||||||||||||||| |||||||||||||| |||||| |||| || |||||||||
Db  86571  CTGGGATTATAGGCGTGAGCCACCCACCCTGGCCAAAAAAGTAAAAAAACTATATAATAAT  86512

Qy    181  AGGGAATAGTTATAGTGTAC---CCCATTTACCCATCACTCAGTTTCAACAGCTGGTGCAC  237
           ||||||| |||||| | || ||   ||||| || || ||| ||||| |||| || |||
Db  86511  ATTAAAGCTTTTTTTAAAGTGACACATGGGGAAATAACCACCCCCTTGATTTCAATGATCAGTCAA  86452

Qy    238  ATATTTATTTCTTCTATACCAGTACCGTACTCTCCCCACTGGGGATTATTTTAAGGCAAAA  297
            |  | ||| | |||||| ||| ||||||| |||||||| |||| |||| | |||||||
Db  86451  TAACTTAAAAATTGTTCCCTGGCAGAAACACAAAATCTTCAAAGCTTGCTTATTGAATA  86392

Qy    298  CCCAGATGACACATTTTTATCCCTAAATACTTTAGATAAAGGTGTTCTTTGAAAAAAA---  352
           | ||||  ||||| ||||||||||||||||  || |||| || || |||||| ||||
Db  86391  GAAGTACTGTACTTTGGCTGGAAGTATTTGTTACTCCCCAGCGACACTTTGGGAGGCAAACCCTGTCTG  86332

Qy    353  -----------TCATAAACCTCAGGACCAGCCTGGCCAACAnGGTGAAAACCCTGTCTG  398
                      ||| | | ||||| ||||||||| |||| |||||| |||| |||
Db  86331  TGGATCACCTGAGGTCAGGAGTTCGAGACCAGCCTGGCCAAAATGGTGAAAACCTCGATCTC  86272

Qy    399  TACTAAAAATACAAAAAATTAGCTTGGCATGGTGGCGCACCTGTAATCCCAGCTACTCA  458
           |||||||||||||||||||||||||||||||||| || |||| |||||||||||||||
Db  86271  TACTAAAAAATACAAAAAAATTAGCCACGTGCGGCGGTGCCTGTAATCCCAGCTACTCT  86212

Qy    459  GGAAGCTGAGGCAGGAGAATCACTTGAATCCGGGAGGCAGAGGTTGCAGTGAGCTGAGAT  518
           ||| ||||||||||||||| | |||| |||||||||| || || ||||||| ||| |||
Db  86211  GGAGGCTGAGGCAGGAGAATTGCTTGAACCCGGGAGGCGGAGGTTGCAGTGAGCCGATAT  86152
```

```
Qy    519 TGCAGTCGAGCCTGGGCGACAGAGACAGA-AATGAAAACTCTGTCTCAAAAACAAACAAAC 577
          |||                          |||  ||||||||||||||||  ||||
Db  86151 CGCACCACTGCACTCCAGCCTGGGCGCGACAGAGTGAGACTCCGTCTTCAAAAAAAAAAAAA 86092

Qy    578 AAA 580
          | |
Db  86091 AGA 86089
```

RESULT 10
US-09-741-154-3/c
Sequence 3, Application US/09741154
Patent No. 6437110
GENERAL INFORMATION:
APPLICANT: BEASLEY, Ellen M. et al
TITLE OF INVENTION: ISOLATED HUMAN KINASE PROTEINS, NUCLEIC
TITLE OF INVENTION: ACID MOLECULES ENCODING HUMAN KINASE PROTEINS, AND USES
TITLE OF INVENTION: THEREOF
FILE REFERENCE: CL001061
CURRENT APPLICATION NUMBER: US/09/741,154
CURRENT FILING DATE: 2000-12-21
NUMBER OF SEQ ID NOS: 4
SOFTWARE: FastSEQ for Windows Version 4.0
SEQ ID NO 3
LENGTH: 16389
TYPE: DNA
ORGANISM: Human
US-09-741-154-3

Query Match   18.1%; Score 180.6; DB 4; Length 16389;
Best Local Similarity 61.5%; Pred. No. 1.9e-38;
Matches 389; Conservative 0; Mismatches 189; Indels 55; Gaps 4;

```
Qy      1 CCTGCCACCATGCCTGGCTAATTTCTTATTTTTAGTAGAGACGAGGTTTTGCCATGTTG 60
          |||  |||||||||||||||||||||    |  ||||||||||||||| |||  ||||||||
Db    870 CCCACCACCATGCCTGGCTAATTTTTGTATTTTTAGTAGAGACGGGGTTTCACCATGTTG 811

Qy     61 ACCAGGCTGGTCTCGAACTCCTAGGTGATCGCCTGCCTCAGCCTCCCAAAGTG 120
          |||||||||||||||||||||| || |||| |||| |||| | ||||||||| 
Db    810 GCCAGGCTGGTCTCGAACTCCTCAGGCGATCTGCCCGCCTCGGCATCCCAAAGTA 751

Qy    121 CTGGGATTATAGGCGTGAGCCGCCGCGCCCAGCCAAC----------- 157
          |||||||||| ||||||||||| ||| ||||| | ||              
Db    750 TTGGGATTACAGGCGTGAGCCACCGCACCCACCTCCCCGTTCTTGAGTGTT 691

Qy    158 -----ATTTTTTAAATACTGAAAAGTAGAGGGAATAGTTATAGTAGTTAC---CCCATTTA 208
               ||| | | ||| |||| |  |       ||  | ||| | || ||   ||||| |
Db    690 TTTCTATTTCTTTCCTATCCAAATCTCACAGTGTTCTTCACATATCACTCTCCTCCCATAGA 631

Qy    209 CCCATCACTCAGTTCAACAGCTGGGTGACATATATTTATTTCTTCTATACCAGTACCGTACT 268
          ||| | ||| ||  | ||  |   | |   ||| |    |  ||| | | ||| || |||
Db    630 CCCCCCCTCGGCCTGTCATAAATCTGCCTAACCCACTACTACTCTCAAATTCCTAGGATCACT 571

Qy    269 CTCCCACTGGGATTATTTTAAGGCAAACC-------------CAGATGACATT 310
          |||   | || |||| | || || |||||              ||| ||| ||
Db    570 GAGACCAGTCTCTCATTTGAGAAGTAGAACACCTGAGCCGCCGGGCAGCGTGGCTCACC 511

Qy    311 TTATCCCTAAAATACTTTAG-----ATAAAGGTGTTCTTTGAAAAAATCATAACCTCAGG 365
          |  | |||| |||| ||  |      | || |||| | ||  ||||||| ||| | |||
Db    510 TGTAATCCCAGCACTTTGGGAGGCCGAGGTGGGTTGGTTCACCTGAGGTCAGGGTTCAG 451

Qy    366 ACCAGGCCTGGCCAACATGGTGAAACCCTGTCTCTGTACTAAAAATACAAAATTAGCTTGGC 425
          ||||||||||||||||| |    |||| || ||| |||||||||||||||| |||||| |||
Db    450 ACCAGGCCTGGCCAACACACCCGTCTCTACTAAAATACAAAAATTAGCTGGG 391

Qy    426 ATGGTCGTCGTGGGCACCTGTAATCCCAGCTACTCAGGAGGCAGGAGAATCACTTGA 485
          |||| | | |||| ||||| |||||||||| ||||||||||| |||| ||| ||||
Db    390 GTGGCGGGCGGGCGCCTATAATCCCAAGTACTCAGGAGGCTGAGGTAGGAGAATGCTTGA 331

Qy    486 ATCCGGGAGGCAGCAGAGATTGCAGTGAGCTGAGATTGCACCACTGCACTCCAGCCTG 545
          |||| |||||| |||  ||||||||||||||| |||| ||||||| | | ||||  |||
Db    330 ATCCAGGAGGCTGAGGTTGCAGTGAGCCAAGATCGCACCACTATAGACCAGTCTGGCAAC 271
```

RESULT 11
US-09-813-817-3/c
Sequence 3, Application US/09813817
Patent No. 6340583
GENERAL INFORMATION:
APPLICANT: YAN, Chunhua et al.
TITLE OF INVENTION: ISOLATED HUMAN KINASE PROTEINS, NUCLEIC
TITLE OF INVENTION: ACID MOLECULES ENCODING HUMAN KINASE PROTEINS, AND USES
TITLE OF INVENTION: THEREOF
FILE REFERENCE: CL001178
CURRENT APPLICATION NUMBER: US/09/813,817
CURRENT FILING DATE: 2001-03-22
NUMBER OF SEQ ID NOS: 4
SOFTWARE: FastSEQ for Windows Version 4.0
SEQ ID NO 3
LENGTH: 59065
TYPE: DNA
ORGANISM: Human
US-09-813-817-3

Query Match   17.5%; Score 174.8; DB 4; Length 59065;
Best Local Similarity 60.4%; Pred. No. 1.1e-36;
Matches 359; Conservative 0; Mismatches 227; Indels 8; Gaps 4;

```
Qy    546 GAAATGAAAACTCTGTCTCAAAAACAAACAAACA 578
          |||| |||| ||| ||||| ||||| |||| |  |
Db    270 AGAGTGAGACTCCGTCTTCAAAAAAAAAAAAAA 238

Qy      4 GCCACCATGCCTGGCTAATTTTCTTATTTTTAGTAGAGACGAGGTTTTGCCATGTTGACC 63
          ||||||| | |||| |||||||  | ||||||||||||| | |||| ||| |||||||| ||
Db  12217 GCCACCACGACCCAGCTAATTTTTGTATTTTTAGTAGAGATGGGGTTTCACCATGTTGGCC 12158

Qy     64 AGGCTGGTCTCGAACTCCTGACCTCAGGTGATCCGCTGCCTCAGCCTCCCAAAGTGCTG 123
          ||| |||||||| | |||||||||||| | |||| |||| |||| |||||||||||||||
Db  12157 AGGATGGTCTCAAATCCTGACCTCA--TGATCTGCCCGCCTCGGCCTCCCAAAGTGCTG 12100

Qy    124 GGATTATAGGCGTGAGCCGCCGCGCCCAGCCAACACATTTTTTAAATACTGAAAAGTAGAGG 183
          | ||||| || ||||||| |||||||||| | ||  |||||  | ||   | | | |||
Db  12099 AGATGATTAGGTGTGAGCCACCGCGCCCAGGTTAAGGATTTTGGTCTTTATCCAGA 12040

Qy    184 GAATAGTTATAGTGTACCCCANTTACCCATCACTCAGTTTCAACAGCTGGGTGACA---TA 240
          ||||  | || ||   | |   ||||| | || |  | |   |   |   | |||    ||
Db  12039 ACACTCGGGAAGTTGTGGAAGTATACAAGGTAGGAGAGACAACACAGAGTCAA 11980

Qy    241 TTTATTTCTTCTATACCAGTACCGTACTCTCCCCACTGGGATTATTTTAAGGCAAACCC 300
          |||   |  |  |   | | |    | | | || | |  |  || || || | |||||
Db  11979 GTTTTTTTTTTTTTTTTTGTTTTTTTTTTAAGATCACTCCTTCCCACCGTGCCAGTGGGTCAC 11920

Qy    301 AGATGACATTTTTATCCCTAAAATACTTTAGATGTGTTTCTTTGAAAAAATCATAACC 360
          | | ||| || |  | |||| ||||| |||| | || | ||  ||||||| ||| |
Db  11919 ACCTGTAATCCCAGCACTTTGGGAGGCCGAGGCAGGTGGA--TCACTTGAGGTCAGGAGT 11862

Qy    361 TCAGGACCAGCGCTGGCCAACATGGTGAAACCCTGTCTGTACTAAAAATACAAAATTAGC 420
          || || ||| | ||||||||||||| ||||||| | ||| |||||||||| |||||||||
Db  11861 TCGAGACTACCTTGGCCAACATGGCAAAACCCCATCTCTACTAAAAATATAAAAATTAGC 11802

Qy    421 TTGGCATGGTGTCGTGGGCACCTGTAATCCCAGCTACTCAGGAGGCTGAGGCAGGAGAATCA 480
          | ||  ||| | | ||| | ||||| |||||||||| |  |||||||||||||||||||| |
Db  11801 TGGGTATGGGGGCGGGCGCCTGTA-TAATCCCAGCTACTTGGGAGGCTGAGGCAGGAGAATTG 11743

Qy    481 CTTGAANTCCGGGAAGCGAGAGAGATTGCAGTGAGCTGAGATTGCAGTGAGCCGACAG 540
          ||| |   ||||||  | | || | ||||||||||| ||| |||| | ||||  |  |
Db  11742 CTTAAACCCGGGAGGCGGAGAAGTTGCAGTGAGCCGAAATTGTGCCACTGCACTCCAGCCTT 11683

Qy    541 AGACAGAAATGAAACTCTGTCTCAAAAACAAACAAACAAACAACAAAACCACTATACA 594
          ||||| | ||| |||| | ||| |||||||||| || || |||| | | |||| ||
Db  11682 GGGGAACAGAGTGAGACTCTCTGTCTAAAACAAAACAAAAACAACAACAAAATCA 11629
```

RESULT 12
US-09-978-197-3/c
Sequence 3, Application US/0978197

; Patent No. 6403353
; GENERAL INFORMATION:
; APPLICANT: YAN, Chunhua et al.
; TITLE OF INVENTION: ISOLATED HUMAN KINASE PROTEINS, NUCLEIC
; TITLE OF INVENTION: ACID MOLECULES ENCODING HUMAN KINASE PROTEINS, AND USES
; TITLE OF INVENTION: THEREOF
; FILE REFERENCE: CL00117BDIV
; CURRENT APPLICATION NUMBER: US/09/978,197
; CURRENT FILING DATE: 2001-10-17
; PRIOR APPLICATION NUMBER: 09/813,817
; PRIOR FILING DATE: 2001-03-22
; NUMBER OF SEQ ID NOS: 4
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 3
; TYPE: DNA
; ORGANISM: Human
; US-09-978-197-3
; LENGTH: 59065

Query Match   17.5%; Score 174.8; DB 4; Length 59065;
Best Local Similarity   60.4%; Pred. No. 1.1e-36;
Matches 359; Conservative   0; Mismatches 227; Indels   8; Gaps   4;

```
Qy      4 GCCACCACCATGCCTGGGCTAATTTTCTTATTTTTAGTAGAGACGAGGTTTTGCCATGTTGACC 63
            |||  ||||  ||| | | | || || ||||||||||||||||||||||||||||  |  ||
Db  12217 GCCACCACGACCAGCTAATTTTTTGTATTTTTAGTAGAGAATGGGGTTTCACCANGTTGGCC 12158

Qy     64 AGGCTGGTCTCGAACTCTTGACCTCAGGTGATCCGCCTCCCAGCCTCCCAAAGTGCTG 123
          |||| ||| |||| ||||| ||| ||    |||| ||||| || |||||||||||| ||
Db  12157 AGGATGGGTCTCAAACTCTCTTGACTTCA--TGATCTGCCTCGCCTCCCAAAGTGCCTG 12100

Qy    124 GGATTATAGGCGTGAGCGCCGCGCCCACCGCCAACACATTTTTTAAATACTGAAAAGTAGAGG 183
            |||||||||  || ||||| |||| |||| |||   |  ||| |||  | |  ||  |
Db  12099 AGATGATAGGCGTGTGAGCGCCACCGCGCCCTGGCCAAGTTAAGGATTTGGTCTTTATCCAAGA 12040

Qy    184 GAATAGTTATAGTGTGTACCCCATTTACCCACTCAGTTTCAACACGTGGTCACA--TA 240
             || | |  | |  ||  |  |  || ||  |||    ||| ||| || |   |
Db  12039 ACACTGGGAAGTGTGGAAGTATACAAGGTAGGAGAGACAACACAGAGTGACACAGTCAA 11980

Qy    241 TTTATTTCTTCTATACCAGTACCGTACTCTCCCCACTGGGGATTAATTATTTTAAGGCAAAACCC 300
            ||| | | | | |    |  ||    | |||| || || |  |  |  |    |||  | |
Db  11979 GTTTTTTTTTGTGTTTTTTTTTTTTAGATACACTCCTTCCACGCCAGTGCACTGGCTCAC 11920

Qy    301 AGATGACATTTTATCCCCTAAATACTTTAGATAAAGGTGTCTTCTTTGAAAAAAATCATAACC 360
           | | | |  || |  | |   | ||    | | |  | || |    |||   | || | |
Db  11919 ACCTGTRAATCCCAGCACTTTGGGAGGCCGAGGCGTGGA--TCACTTGAGGTCAGGAGT 11862

Qy    361 TCAGGACCAGCCTGGGCAACACATGGTGAAAACCCTGTCTCTGTACTTTAAAAATACAAAAATTAGC 420
          |||| |||||| ||||||||||| |||  ||||||| ||||  ||||  |||||| |||||||||||
Db  11861 TCGAGACTAGCTTGGCCAACATGGCAAAACCCCATCTCTACTAAAAATATAAAAATTAGC 11802

Qy    421 TTGGCATGGTCGTGGGCACCTGTAATCCCAGCTACTCAGGAGGCTGAGGCAGGAGAATCA 480
           |||| ||| | ||| | |||||||||||||||||| | ||||||||||||||||||||
Db  11801 TGGGTATGGGGGCGGGTGCCT-TAATCCCAGCTACTTGGGAGGCTGAGGCAGGAGAATTG 11743

Qy    481 CTTGAATCCGGGAGGCAGAGAGATTGCAGTGAGCTGAGATCGAGACCCTGGGCGACAG 540
          ||| ||  |||||||| ||||  ||||||||||| ||||| | |  |||||| ||| |
Db  11742 CTTAAACCGGGAGGCGGAGGTTGCAGTGAGCCGAGATTGTGCCACTGCACTCCAGCCTT 11683

Qy    541 AGACAGAAATGAAAACTCTGTCTGTCTCAAAAACAAACAAACAAACAAAAAAAACACTATACA 594
          ||||||| |   ||| ||||||| || |||||||||  |||| || |||||    | ||| |
Db  11682 GGGGAACAGAGTCAGACTCTGTCTCTAAAAACAAAACAAAACAACAAAAAATCA 11629
```

; FILE REFERENCE: 7853-138
; CURRENT APPLICATION NUMBER: US/09/268,992
; CURRENT FILING DATE: 1999-03-16
; EARLIER APPLICATION NUMBER: 09/236,134
; EARLIER FILING DATE: 1999-01-22
; EARLIER APPLICATION NUMBER: 60/106,056
; EARLIER FILING DATE: 1999-01-28
; EARLIER APPLICATION NUMBER: 60/088,312
; EARLIER FILING DATE: 1998-06-05
; EARLIER APPLICATION NUMBER: 60/078,044
; EARLIER FILING DATE: 1998-03-16
; NUMBER OF SEQ ID NOS: 84
; SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 7
; LENGTH: 72604
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: modified_base
; LOCATION: all n positions
; OTHER INFORMATION: n=a, c, g, or t
; US-09-268-992-7

Query Match   17.1%; Score 171.4; DB 4; Length 72604;
Best Local Similarity   62.4%; Pred. No. 9.4e-36;
Matches 386; Conservative   1; Mismatches 192; Indels 40; Gaps   6;

```
Qy      5 CCACCATGCCTGGGCTAATTTCTT--ATTTTTAGTAGAGACGAGGTTTTTGCCATGTTGAC 62
          ||||||||||||| | ||| ||    |||||||||||||||| |||| |||| |||||| |
Db  45015 CCGCCATGCCTGGCCTGGCCTAATTTTTTGTATTTTTAGTAGAGACAGGGTTTTACCGTGTTGCC 44956

Qy     63 CAGGCTGGTCTCGAACTCTTGACCTCAGGTGATCCGCCTCAGCCTCCCAAAGTGCT 122
          ||||||||||||||||||  || |||||| ||||| | ||| ||||||||||||| |
Db  44955 CAGGCTGGTCTCGAACTCCTGAGCTCAGGKAATCCACATGCCTCAGCCTCCCAAAGTGCCT 44896

Qy    123 -GGGATTATAGGCGTGAGCGCGCGCCACCGCGCCCAGCCACATTTTTTAAATACTGAAAAGTAGA 181
           |||||  |||| ||||||  | || || || |||  || ||||| ||    | ||||  |||
Db  44895 AGGGATAACAGGGGTGAGCCACTGCACCCGCCCTGGACCATTTCTTATTGCAAATTTTA 44836

Qy    182 GGGAATAGTTATAGTGTGTACCCCATTTACCCATCACTCAGTTTCAACAGCTGTGACATAT 241
           | | | | |  | |  || || || | ||    ||  || |||  || |  |  | |||
Db  44835 AAGTCAACTTACAGTGAGTGCATTTCAAATCAATAATAATTCTGGGCAACGCTAGTATGAGGC 44776

Qy    242 TTATTTCTTCTATACCAGTACCGTACTCT-----CCCCACTGGGGATTAATTATTTTAAGGCAA 295
           ||| | || | | | | | | | || ||     | |  |||   | ||   | ||  | |
Db  44775 AATTAACTTTAAGATAANTTTCTTAAACCCTCTATACCTGACTTTGCCTTTGCCTATGGAA 44716

Qy    296 AACCCAGAGATGACATTTTATC-------------CCTAAATACTTTAGATA 332
          | | | || || || ||  | |             |||  ||||  || |
Db  44715 ACAGGAACTGAAATTTCTCCTGGTTTACATTGCCCACCAATAGCCTAAAAAGGTCAAGTT 44656

Qy    333 AAGGTGTTCTTTGAAAAAAAATCATAACCTCAGG------------ACCAGCCTGGGCCAACATGGT 385
           | | || |   |  |  || |||||| ||| ||            |||||| |||||||||| |||
Db  44655 TTCACTTGGGAGGCCAAGGTGTGTCGGATCACCTGAGGTCACCAGCGCTGGGCCAACATGGT 44596

Qy    386 GAAACCCTGTCTGTACTAAAAATACAAAAATTACCTTGGCATGGTGGTCGTGGGCACCTGTAA 445
          ||||||||||||| ||||||||||||||||||| || || | ||||||| |||||| |||||
Db  44595 AAAAACCCTGTCTCTACTAAAAATACAAAAATTAGCTAGCCGTGGTGGTGGGCGCCTGTAA 44536

Qy    446 TCCCAGCTACTCAGGAAGCTGAGGCAGGAGAATCACTTGAATCCGGGAGGCAGAGGTTGC 505
          |||||||||||||||| ||||||||||||||||||||||| ||||||||| ||||| |||
Db  44535 TCCCAGCTACTCAGGAGGCTGAGGCAGGAGAATCACTTGAACCCGGGAGGTGGAGGCTGC 44476

Qy    506 AGTGAGCTGAGATCGAGCCGTGGGCGACAGAG-ACAGAAATGAAAACTCTGTCTCA 564
          ||| ||| ||||| | || | |||| | ||||  |||||||||||| |||| ||||
Db  44475 AGTAAGCCAAGATGGCACCACTGCCCTCTAGCCAGGGCAAAAGAGCCAGACCCTGTGTCTCA 44416

Qy    565 AAAACAAACAAACAAAAAA 583
          |  | |||| ||| ||||
Db  44415 TGAAAAAAAATAAAGAAAA 44397
```

RESULT 13
US-09-268-992-7/c
Sequence 7, Application US/09268992
; Patent No. 6342351
; GENERAL INFORMATION:
; APPLICANT: Chen, H.
; APPLICANT: Freimer, N.
; TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR DIAGNOSING
; TITLE OF INVENTION: AND TREATING CHROMOSOME-18p RELATED DISORDERS

RESULT 14

US-09-657-474-7/c
; Sequence 7, Application US/0965747474
; Patent No. 6399762
; GENERAL INFORMATION:
; APPLICANT: Freimer, N.
; APPLICANT: Chen, H.
; TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR DIAGNOSING
; TITLE OF INVENTION: AND TREATING CHROMOSOME-18p RELATED DISORDERS
; FILE REFERENCE: 7853-138
; CURRENT APPLICATION NUMBER: US/09/657,474
; CURRENT FILING DATE: 2000-09-07
; PRIOR APPLICATION NUMBER: 09/268,992
; PRIOR FILING DATE: 1999-03-16
; PRIOR APPLICATION NUMBER: 09/236,134
; PRIOR FILING DATE: 1999-01-22
; PRIOR APPLICATION NUMBER: 60/106,056
; PRIOR FILING DATE: 1998-10-28
; PRIOR APPLICATION NUMBER: 60/088,312
; PRIOR FILING DATE: 1998-06-05
; PRIOR APPLICATION NUMBER: 60/078,044
; PRIOR FILING DATE: 1998-03-16
; NUMBER OF SEQ ID NOS: 84
; SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 7
; LENGTH: 72604
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: modified_base
; LOCATION: all n positions
; OTHER INFORMATION: n=a, c, g, or t
US-09-657-474-7

Query Match　　17.1%;　Score 171.4;　DB 4;　Length 72604;
Best Local Similarity　62.4%;　Pred. No. 9.4e-36;
Matches 386;　Conservative　1;　Mismatches 192;　Indels 40;　Gaps 6;

```
Qy      5 CCACCATGCCTGGCTAANTTTCTT--ATTTTTAGTAGAGACGGAGGTTTTGCCATGGTTGAC 62
          ||||||||||||||||||  ||||||   | ||||||||||||||||||||||||||||||||
Db  45015 CCGCCCATGCCTGGCCTAATTTTTTTTTGTATTTTAGTACAGACAGGGTTTTCACCGTGTTGCC 44956

Qy     63 CAGGCTGGTCTCGAACTCTTGACCTCAGGTGATCGCGCCTGCCTCAGCCCTCCCAAAGTGCT 122
          ||||||||||||||||||| ||||||:  ||||| ||| |||||||||||| |||||||||||
Db  44955 CAGGCTGGTCTCCGAACTCCTGAGCTCAGGKAATCCACATGCCTCAGCCCTCCCAAAGTGCT 44896

Qy    123 -GGGATTATAGGCGTGAGCCGCCGCACCCAGCCACAACATTTTTAAATACTGAAAAGTAGA 181
           |||||||||||||||| ||||||||||||||||| |||| ||||| |||||||||||||||
Db  44895 AGGGATAACAGGGGTGAGCCACTGCACCCGCCCTGGACCATTTCTTAATTGCAAAATTTTA 44836

Qy    182 GGGAATAGTTATAGTGTACCCCATTACCCATCAGTTTCAAACAGCTGGTGTGACACTAT 241
          ||||||| ||||||||||| |||||| ||||||||| |||  |||||||| ||||||||
Db  44835 AAGTCAACTTACAGTAGTCATTTCAAATCAATAATAATTCTGGGCAGCTAGTGATGAGGC 44776

Qy    242 TTATTTCTTCTATACCAGTACCGTACTCT----CCCCACTGGGGATTATATTTTAAGGCAA 295
          ||| || |  ||||||||| |||| ||||    ||||  ||| ||||||||||||||||
Db  44775 AATTAACTTTAAGATAATTTCTTAAACCCTCTTATACCTGACTTGGCCTTTGCCTTATGGAA 44716

Qy    296 AACCCAGAATGACANTTTTATC----------CCTAAATACNTTAGNATA 332
          |  ||||||  || | ||| |||           ||||| ||  |||||||
Db  44715 ACAGGATCTGAAAATTCTCCTGGTTTACATTGCCCACCAAATAGGTCAAGTT 44656

Qy    333 AAGGTGTTGTCTTTGAAAAAAATCATAACCTCAGG--------ACCAGCCTGGGCCAACATGGT 385
           | || |||  || |||||||||||||||||||||         |||| ||||| |||||||||
Db  44655 TTCACTTTGGGAGGCCAAGGTGTGTGGAATCACCTGAGGTCACCAGCCTGGGCCAACATGGT 44596

Qy    386 GAAACCCTGTCTCTGTTACTAAAAAATACAAAAAAATTAGCTGTGGTGCCACCTGTAA 445
          |  |||||||||||||| |||| ||||| || ||| ||||| |||||||| ||||||
Db  44595 AAAACCCTGTCTCTACTAAAAAATACAAAAAATTAGCTGGGCGTGGTGGTTTGGGCCTGTAA 44536

Qy    446 TCCCAGCTACTCAGGAGGCTGAGGCAGGAGAATCACTTGAATCCGGGAAGCAGGAGATTGC 505
          |||| | |||| ||||||| |||| |||||||||||||||||||||||||||||| |  ||
Db  44535 TCCCAGCTACTCAGGAGGCTGAGGCAGGAGAATCACTTGAACCCAGGAGGTGGAGGCTGC 44476
```

US-09-798-096-10/c
; Sequence 10, Application US/09798096
; Patent No. 6399378
; GENERAL INFORMATION:
; APPLICANT: Andrew T. Watt
; APPLICANT: Donna T. Ward
; TITLE OF INVENTION: ANTISENSE MODULATION OF RECQL2 EXPRESSION
; FILE REFERENCE: RTS-0207
; CURRENT APPLICATION NUMBER: US/09/798,096
; CURRENT FILING DATE: 2001-03-01
; NUMBER OF SEQ ID NOS: 89
; SEQ ID NO 10
; LENGTH: 99500
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; FEATURE:
US-09-798-096-10

RESULT 15

Query Match　17.0%;　Score 169.6;　DB 4;　Length 99500;
Best Local Similarity　61.2%;　Pred. No. 3.2e-35;
Matches 361;　Conservative　0;　Mismatches 194;　Indels 35;　Gaps 4;

```
Qy      5 CCACCATGCCTGGCTAATTTTTCT--TATTTTTAGTAGAGACGGAGGTTTTTGCCATGGTTGAC 62
          |||||||||||||||||  |||||||   | ||||||||||||| |||||||||||||||||||
Db  28687 CCACCACGCCCCCACTAATTTTTTGTGTATTTTTGGTGGAGATGGGGGTTTCACCATGTTTCC 28628

Qy     63 CAGGCTGGTCTCGAACTCCTTGACCTCAGGTGATCGCGCCTGCCTCGCCCTCCCAAAGTGCT 122
          ||||||||||||||||||||  |||||| |||| ||| ||||||||| ||||||||||||||
Db  28627 CAGGCTGGTCTCGAACTCCTGTGGTCCTCCAAGGATCCTGCCTCAGCCTCCCAAAGTGCT 28568

Qy    123 GGGATTATAGGCGTGAGCCGCCGCACCCAGCCAACAATTTTAAATACTGAAAAGTAGAG 182
          ||||||||||| |||||||||||||||||| |||||||  ||| | |||||| || |||
Db  28567 GGGATTGCAGGTGTGAGCCACCGCGCCCGGCCTCCCT------GATTCTTTTTATGGAG 28514

Qy    183 GGAATAGTTATAGTGTACCCCATTTACCCATCACTCAGTTTCAACAGCTGGTGTGACACATT 242
          ||||| || | ||  |||||| ||||  |||  ||||||| |  ||||||| |||| ||||
Db  28513 CATATTTTAAGCCCCTTTAATCTCTGCATTTCCTAGGCACTTCAGCTTTGGAATCCAGTT 28454

Qy    243 TATTTCTTTCTATACCAGTACCGTACTCTCCCCCACTGGGGATTATTTTTAAGGCAAAACCCAG 302
          | | |||||||||||  |||||  |  || |||| |||||||||||||||||||||||||||
Db  28453 TAAAAGT---------------------ATTGATCTTGGCTGGGCACGGTGGCTC 28420

Qy    303 ATGACANTTTATCCCTAAAATACTTTAGATAAAAGGTGTTCTTTGAAAAAAATCATAACCTC 362
          |||||| | ||||   |    | |||| |||||||||  | || |||||||||||||||||
Db  28419 ACGCCTGTAATCCCAGCACTTTGGAAGGCCGAGGCGGGCGGATCATGAGGTCAGGAGATC 28360

Qy    363 AGGACCAGCCTGGGCCAACATGGTGAAACCGTCTGTACTAAAAATACAAAAAATTAGCTTT 422
          |||| |||||||| |||||||| |||| |||  ||| | ||||||||||||||||| |||
Db  28359 AAGACCAACCTGGCCAATAIGGTGAAACCCCGTCTCTACTAAAAATACAAAAAATTAGCTG 28301

Qy    423 GGCATGGTGGCGCATGCCTGTAATCCCAGCTACTCAGGGAGGCTGAGGCAGGAGAATCGCAGT 482
          ||||| || ||| ||||||||||||||||||||||||| ||||||||||||||||||||||
Db  28300 GGCATGGTGGCGCGCGCCTGTAATCCCAGCTACTTGGGAGGCTGAGGCAGGAGAATCACT 28241

Qy    483 TGAATCGGGAGGCAGAGGATTGCAGTGAGCTGAGATTGCGACCTCGGGGCGACAGAG 542
          |||||| ||||| | |||  ||||||| |||||||||| ||| |||||||||  |||
Db  28240 TGAACCCGGGAGGCAGAGGTTGCAGTGAACCGAGATTGTGCCACTGCACTCCAGCCTGGT 28181

Qy    543 ACAGAGAATGAAACTCTGTCTCAAAAACAAAACAAAACAAAACACAGTATA 592
          | ||||| |||||| ||| ||||||||||||||||||||||||||||||||
Db  28180 GACAGAGCTAGACTCTTGTCTCAAAAAATAAAATAAAATAAAAATAAA 28131
```

Qy   506 AGTGAGCTGAGATTGCAGTGCAGCCGTGGGGCGACAGAG-ACAGAGAAATGAAAACTCTGTCTCA 564
Db 44475 AGTAAGCCAAGAATGGCACCACTGCCCTCTCAGCCAGGGCAAAAGAGCCAGACCCTGTCTCA 44416

Qy   565 AAAACAAAACAAAACAAAAAA 583
Db 44415 TGAAAAAAAATAAAGAAAA 44397

Search completed: March 29, 2003, 22:40:54
Job time : 1074.43 secs

OM nucleic - nucleic search, using sw model
          GenCore version 5.1.4_p5_4578
          Copyright (c) 1993 - 2003 Compugen Ltd.

Run on: March 29, 2003, 21:53:16 ; Search time 105.307 seconds
                              (without alignments)
                              8080.893 Million cell updates/sec

Title: US-09-691-219-3_COPY_1_1000
Perfect score: 1000
Sequence: 1 cctgccacatgcctggcta.........caagaggcgctgcacctctga 1000
Scoring table: IDENTITY_NUC
               Gapop 10.0 , Gapext 1.0
Searched: 574371 seqs, 425486471 residues
Total number of hits satisfying chosen parameters: 1148742
Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
               Maximum Match 100%
               Listing first 45 summaries

Database :   Published_Applications_NA:*
     1: /cgn2_6/ptodata/2/pubpna/US07_PUBCOMB.seq:*
     2: /cgn2_6/ptodata/2/pubpna/US07_NEW_PUB.seq:*
     3: /cgn2_6/ptodata/2/pubpna/US06_NEW_PUB.seq:*
     4: /cgn2_6/ptodata/2/pubpna/US06_PUBCOMB.seq:*
     5: /cgn2_6/ptodata/2/pubpna/US07_NEW_PUB.seq:*
     6: /cgn2_6/ptodata/2/pubpna/PCTUS_PUBCOMB.seq:*
     7: /cgn2_6/ptodata/2/pubpna/US08_NEW_PUB.seq:*
     8: /cgn2_6/ptodata/2/pubpna/US08_PUBCOMB.seq:*
     9: /cgn2_6/ptodata/2/pubpna/US09_NEW_PUB.seq:*
    10: /cgn2_6/ptodata/2/pubpna/US09_PUBCOMB.seq:*
    11: /cgn2_6/ptodata/2/pubpna/US10_NEW_PUB.seq:*
    12: /cgn2_6/ptodata/2/pubpna/US10_PUBCOMB.seq:*
    13: /cgn2_6/ptodata/2/pubpna/US60_NEW_PUB.seq:*
    14: /cgn2_6/ptodata/2/pubpna/US60_PUBCOMB.seq:*

SUMMARIES

	Result No.	Score	% Query Match Length	DB	ID	Description
c	1	202	20.2	32192	9	US-10-092-154-1416	Sequence 1416, Ap
c	2	202	20.2	32192	10	US-09-764-847-1416	Sequence 1416, Ap
c	3	201.8	20.2	11538	10	US-09-764-869-1598	Sequence 1598, Ap
c	4	201.4	20.1	7703	10	US-09-764-887-415	Sequence 415, App
c	5	201.4	20.1	7703	10	US-09-764-887-456	Sequence 456, App
c	6	199.2	19.9	65608	9	US-09-954-531-180	Sequence 180, App
c	7	199.2	19.9	65608	10	US-09-962-436-292	Sequence 292, App
c	8	199.2	19.9	65608	10	US-09-962-832-119	Sequence 119, App
c	9	194.8	19.5	9887	9	US-10-000-639-7	Sequence 7, Appli
c	10	194.2	19.9	65608	10	US-09-801-574-61	Sequence 61, Appl
c	11	191.2	19.1	53542	10	US-09-801-574-61	Sequence 61, Appli
c	12	191.2	19.1	465237	10	US-09-933-267A-1	Sequence 1, Appl
c	13	189.6	19.0	4150	12	US-10-016-725-15	Sequence 15, Appl
c	14	189.4	18.9	65608	9	US-09-954-531-180	Sequence 180, App
c	15	189	18.9	65608	10	US-09-962-436-292	Sequence 292, App
c	16	189	18.9	65608	10	US-09-962-832-119	Sequence 119, App
c	17	188.2	18.9	15016	10	US-09-880-107-3783	Sequence 3783, Ap
c	18	187	18.7	57130	10	US-09-835-081-3	Sequence 3, Appli
c	19	186.4	18.6	25423	9	US-10-072-349-193	Sequence 193, Appl
c	20	184.8	18.5	25424	9	US-10-072-349-194	Sequence 194, App
c	21	184.8	18.5	25424	10	US-09-764-855-194	Sequence 194, App
c	22	184.6	18.5	25423	9	US-10-072-349-193	Sequence 193, Ap
c	23	184.6	18.5	25423	10	US-09-764-855-193	Sequence 193, App
c	24	184.6	18.5	25424	9	US-10-072-349-194	Sequence 194, App
c	25	184.6	18.5	25424	10	US-09-764-855-194	Sequence 194, App
c	26	184	18.4	15266	10	US-09-764-877-3797	Sequence 3797, Ap
c	27	184	18.4	15266	10	US-09-764-877-3797	Sequence 3797, Ap
c	28	184	18.4	15271	10	US-09-764-877-3798	Sequence 3798, Ap
c	29	184	18.4	15271	10	US-09-764-877-3798	Sequence 3798, Ap
c	30	184	18.4	174493	10	US-10-238-709-3	Sequence 3, Appli
c	31	184	18.4	174493	9	US-10-238-709-3	Sequence 3, Appli
c	32	183.8	18.4	31871	10	US-09-764-884-1403	Sequence 1403, Ap
c	33	183.8	18.4	31871	9	US-10-092-154-1403	Sequence 1403, Ap
c	34	180.6	18.1	174424	10	US-09-967-768A-314	Sequence 314, App
c	35	180	18.0	7703	10	US-09-764-887-415	Sequence 415, App
c	36	180	18.0	7703	10	US-09-764-887-456	Sequence 456, App
c	37	179	17.9	1086	10	US-09-764-869-2140	Sequence 2140, Ap
c	38	178	17.8	19696	9	US-10-091-572-874	Sequence 874, App
c	39	177.4	17.7	1575	10	US-09-822-846-257	Sequence 257, Appl
c	40	177.2	17.7	41907	10	US-09-967-013-5	Sequence 5, Appli
c	41	177	17.7	32249	10	US-09-764-869-2326	Sequence 2326, Ap
c	42	176.6	17.7	32249	10	US-09-764-869-2314	Sequence 2314, Ap
c	43	175.8	17.6	116592	10	US-09-818-512-3	Sequence 3, Appli
c	44	175.6	17.6	14152	10	US-09-764-869-1573	Sequence 1573, Ap
c	45	175.2	17.5	30568	10	US-09-764-877-3851	Sequence 3851, Ap

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

ALIGNMENTS

RESULT 1
US-10-092-154-1416/c
Sequence 1416, Application US/10092154
; Publication No. US20030054375A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
; FILE REFERENCE: PC009C1
; CURRENT APPLICATION NUMBER: US/10/092,154
; CURRENT FILING DATE: 2002-03-07
; NUMBER OF SEQ ID NOS: 2003
; PRIOR Application removed - See File Wrapper or Palm
; SOFTWARE: Patentin Ver. 2.0
; SEQ ID NO 1416
; LENGTH: 32192
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-092-154-1416

Query Match   20.2%;   Score 202;   DB 9;   Length 32192;
Best Local Similarity   61.68%;
Matches 382; Conservative 61; Mismatches 220; Indels 18; Gaps 3;

```
Qy      1 CCTGCCACCATGCCTGGCTAATTTTCTTATTTTATTTTTAGTAGAGACGAGGTTTGCCATGTTG 60
          |||||||  |||| |||   ||||||| |||| |||||| |||||| | ||| |||||||| |||
Db  26170 CCCGCCACCCGCGCCCAGCTAATTTTTTTGTATTTTACTAGAGAATGGGGTTTTGCCATATTG 26111

Qy     61 ACCAGGCTGGTCTCGAACTCTTGACCTCAGGTGATCCGCCTGCCTCAGCCTCCCAAAGTG 120
          | ||||||||||| ||||| ||||||||||||||||| |||||||| |||||||||||||
Db  26110 GCCAGGCTGGTCTTCAAACTCCTGACCTCAGGTGATCCACCTGCCTCGGCCTCCCAAAGTG 26051

Qy    121 CTGGGATTATAGGCGTGAGCGCCGCGACCAGCCAACATGTTTTTTAAATACNTGAAAAGTAG 180
          |||| ||||| ||| ||||| ||||| ||| |       |||| | ||| | | |||||
Db  26050 CTGGAATTACAGGCATGAGCCACCGTGCCCGGACTTTATTGAAATTAAAAACTTT 25991

Qy    181 AGGGAATAGTTATATAGTTGTACCCCCATTTACCCATCACTCAGTTTCAACAGCTGTGGTGACATA 240
          | |||| | | | |    |||||   || | |  |  |  ||  |||| | || | | | |||||
Db  25990 TGTG-----------CTGTAAATGAAACTATCAATTAGCTGAGAAGACAACATATGTTATACA 25940

Qy    241 TTTATTTCTTCTTATACCAGTAGCACCGTGACTCCTCCCCACTGGGATTATTTTAAGGCAAAAACC 300
          ||  ||| | |  |  ||| |  | |  |   || | |  ||| | || | | |||| |
Db  25939 CAATCTTACAACTCAATAATTCAATAATTTAAAAAACCCTACTAAGCCGGGGCGCGGGCAGTGGC 25880
```

RESULT 2
US-09-764-847-1416/c
Sequence 1416, Application US/09764847
Patent No. US20020132767A1
GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
; FILE REFERENCE: PC009
; CURRENT APPLICATION NUMBER: US/09/764,847
; CURRENT FILING DATE: 2001-01-17
; Prior application data removed - consult PALM or file wrapper
; NUMBER OF SEQ ID NOS: 2003
; SOFTWARE: Patentin Ver. 2.0
; SEQ ID NO 1416
; LENGTH: 32192
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-764-847-1416

Query Match   20.2%;   Score 202;   DB 10;   Length 32192;
Best Local Similarity   61.6%;   Pred. No. 2.6e-48;
Matches   382;   Conservative   0;   Mismatches   220;   Indels   18;   Gaps   3;

```
Qy    301 AGATGACATTTTATCCCTAAATACTTTAGATAAAGGTGTTCTTTTGAAAAAAATCATAACC 360
          ||| |||||
Db  25879 TCACACCTGTAATCCCAGCACTTTGGGAGGCCGAGGCGGGCGGATCACGAGGTCAGGAGA 25820

Qy    361 TCAGGACCAGCCTGGCCAACATGGTGAAAACCCTGTCTCTGTACTAAAATAC-AAAAATTAG 419
          || |||| ||| |||||  |||  ||||||| ||  | || |||||||| ||||||  |||||||||
Db  25819 TCGAGACGATCCTGGCTAACACAGTGAAACCCCGTCTCTACTAAAAATGCAAAAATTAG 25760

Qy    420 CTTGGCATGGTGGCGGGCACCTGTAATCCCAGCTACTCAGGAGGCTGAGGCAGGAGAATC 479
          ||| |   |||||| |||||||||||||||||||||| |  |||||||||||||| |||
Db  25759 GCGTGAACCCCGGGAGGCGGAGGTTGCAGTGCCCAGCTACTCCGGGAAGCTGAGGCGAGGAATG 25700

Qy    480 ACTTGAATCCGGAAGCAGAGATTGCAGTGAGCTGAGAT------TGCAGTCGAGCCT 531
          ||| |||||||||| | |||||| |||||||| ||||||      |||| ||||||
Db  25699 GCGTGAACCCCGGGAGGCGGAGGTTGCAGTGAGCCGAGATCGTGCCACTCCAGCCC 25640

Qy    532 GGGCGACAGGAGCAGGAAATGAAACTCTGTCTCAAAAACAAACAAACAAAAAACCACTAT 591
          |||||||| ||| || ||| ||   ||  | | |||||||||||||||||||||||||
Db  25639 GGGCGACAGGAGACTCCATCTCAAAACAAACAAACAAACAAAAAAAACCCT 25580

Qy    592 ACATAAAAAATGAACAAATGAT 611
          || ||||  | ||| |||
Db  25579 ACTAAAAATATATAAAGGAT 25560
```

RESULT 3
US-09-764-869-1598
Sequence 1598, Application US/09764869
Patent No. US20020061521A1
GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
; FILE REFERENCE: PC007
; CURRENT APPLICATION NUMBER: US/09/764,869
; CURRENT FILING DATE: 2001-01-17
; Prior application data removed - refer to PALM or file wrapper
; NUMBER OF SEQ ID NOS: 2442
; SOFTWARE: Patentin Ver. 2.0
; SEQ ID NO 1598
; LENGTH: 11538
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-764-869-1598

Query Match   20.2%;   Score 201.8;   DB 10;   Length 11538;
Best Local Similarity   63.9%;   Pred. No. 1.6e-48;
Matches   392;   Conservative   0;   Mismatches   202;   Indels   19;   Gaps   5;

```
Qy    361 TCAGGACCAGCCTGGCCAACATGGTGAAAACCCTGTCTCTGTACTAAAATAC-AAAAATTAG 419
          || |||| ||| |||||  |||  ||||||| ||  | || |||||||| ||||||  |||||||||
Db  25819 TCGAGACGATCCTGGCTAACACAGTGAAACCCCGTCTCTACTAAAAATGCAAAAATTAG 25760

Qy    420 CTTGGCATGGTGGCGGGCACCTGTAATCCCAGCTACTCAGGAGGCTGAGGCAGGAGAATC 479
          ||| |   |||||| |||||||||||||||||||||| |  |||||||||||||| |||
Db  25759 GCGTGAACCCCGGGAGGCGGAGGTTGCAGTGCCCAGCTACTCCGGGAAGCTGAGGCGAGGAATG 25700

Qy    480 ACTTGAATCCGGAAGCAGAGATTGCAGTGAGCTGAGAT------TGCAGTCGAGCCT 531
          ||| |||||||||| | |||||| |||||||| ||||||      |||| ||||||
Db  25699 GCGTGAACCCCGGGAGGCGGAGGTTGCAGTGAGCCGAGATCGTGCCACTCCAGCCC 25640

Qy    532 GGGCGACAGGAGCAGGAAATGAAACTCTGTCTCAAAAACAAACAAACAAAAAACCACTAT 591
          |||||||| ||| || ||| ||   ||  | | |||||||||||||||||||||||||
Db  25639 GGGCGACAGGAGACTCCATCTCAAAACAAACAAACAAACAAAAAAAACCCT 25580

Qy    592 ACATAAAAAATGAACAAATGAT 611
          || ||||  | ||| |||
Db  25579 ACTAAAAATATATAAAGGAT 25560

Qy      5 CCACCATGCCTGGCTAATTTCTTTATTTTTAGTAGAGACGAGGTTTGCCATGTTGACCA 64
          ||||||||||| |||||||||||||||  ||||||||  |||||||||||| |||||||
Db   4365 CCACAACGCCCACGTAATTTTTTGTACTTTTAGTAGAGA-GGGGTTTTACCATATTAACCA 4423

Qy     65 GGCTGGTCTCGAACTCCTGACCTCAGGTGATCCGCCCGCCTCAGCCTCCCAAAGTGCTGG 124
          |||||||| |||||||| |||||||||||||  ||||||| ||||||||||||||||||
Db   4424 GGCTGGTCTTGAACTCCTGACCTCAAGTGATCCACCCGCCTCAGCCTCCCAAAATGCTCA 4483

Qy    125 GATTATAGGCGTGAGCCCGCGACCACCCAGCCAACATTTTTAAATACTAAAAGTAGAGGG 184
          | ||| |   | || | || | | |||||||||  ||| || ||   |   ||||
Db   4484 GACTACAGGTGTGAGCCACCGTGCCCAGCC--CAACCATTGTTTTGTTGTTACAGGGATAT 4541

Qy    185 AATAGTTATAGGTATACCCCATTTACCCATCACTCAGTTTCAACAGCTGGTTGACATATTTA 244
          |||||| ||||| ||  ||||  ||| |  ||||| | ||||||||||||||||||
Db   4542 AATGGTGAACTAGACAGCCAAGTTCCCTGCCCCTCATTT------CTATTCTA 4587

Qy    245 TTTCTTCTATACCGTACCGTCTCCCCACTGGGATTATTTTAAGGCAAACCCAGAT 304
          ||| |   |||  |||| | |||||||||||||||| |||||| |||| ||||
Db   4588 CTTCGAGAGGTAGACAATAAAATAGTTTTTTCAGGAAGGCACAGTGGGCTCA 4647

Qy    305 GACATTTTATCCC-TAAATACTTTAGATAAAAGGTGTTCTTTTGAAAAAAATCATAACCTCA 363
          |||| ||  |   |   | |||||||||||  |||||| ||  |||||||||||
Db   4648 CATCTGTAATCCCAGCACTTTGGGAGGCCGAGGCGGGAGGATTGCTTGAGCCCAGGAGTTTG 4707

Qy    364 GGACCAGCCTGGCCAACATGGTGAAACCCGTGTCTGTTACTAAAAATACAAAAAATTAGCC 423
          ||| |||| ||||||||| |||  ||||| ||||| | |  |||||||||  | ||||
Db   4708 AGACCAGTCTGACCAACATGGTGAAACCCTGTCTCTACTAAAAATACAAAAAAATTAGCTAG 4767

Qy    424 GCATGGTGCGTGGGCGCCCTGTAATCCCAGCTACTCAGGAGGCTGAGGCAGGAGAATCACTT 483
```

```
Db  4768 GTGTGATGGTGTGCGGCCTGTAATCCCAGCTACTCGGGAGGCTGAGGCAGGAGAATTGATT 4827
              ||   |||    | |||  | |||||||||||||||||||||||||||||||||||||||  ||
Qy   484 GAATCCGGGAAGCAGAGAGATTGCAGTGAGCCGAGATTGCAGTGAGCCGAGATCGCAGAGA 543
         |||| ||||   |||||||    ||||| | ||| |   ||||||||||||| |||| |||||
Db  4828 GAACCCGGGAGGCGAACAGGTTGCAGTGAGCTGAGATCACACCACTGCACTCCAGCCTGGC 4887

Qy   544 CAGA-AATGAAACTCTGTCTCAAAAAACAAAACAAAACAAAACACTATACATAAAAAATG 602
         ||||  || |  |||  |||||| |  ||  |||||||||||||||||| |||||||
Db  4888 GACAGAGTGAGACTCTGTCTCCAACAAAACAAAACAAAAATTCATATAAAAATACACAGG 4947

Qy   603 AACAATGATGCCA 615
         | ||| |||
Db  4948 TAAAATGTTTCAA 4960
```

```
Query Match         20.1%;    Score 201.4;  DB 10;   Length 7703;
Best Local Similarity  63.1%;  Pred. No. 1.7e-48;
Matches  368;  Conservative   0;  Mismatches 196;  Indels  19;  Gaps  3;

Qy     3 TGCCACCATGCCCTGGCCTAATTTTTCTTAATTTTTTAGTAGAGACGAGGTTTTGCCATGTTGAC 62
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  3644 TGCCACCACGCCCTGGCTAATTTTTGTATTTTTCTAGTAGAGACGGGGTTTCACCATGTTGGC 3703

Qy    63 CAGGCTGGTCTCGAACTCTTGACCTCAGGTGATCCGCCTCAGCCTCCCAAAGTGCT 122
         |||||||||||||||| |||||| |||||||||  ||||||  ||||||||||||||
Db  3704 CAGGCTGGTCTTGAACTCCTGGACCTCAAGTGATCACCACACCCGGCCTCCCAAAGTGCT 3763

Qy   123 GGGATTATAGGCGTGAGCCGCCGCCACCCAGCCAACACATTTTTTAAATACTGGAAAAGTAGAG 182
         ||||||||||||||||||||||  | ||||| |||||||| |||   |||||||||||||||||
Db  3764 GGGATTACAGGCATGAGCCACCGCGCCCGGCCTCAAGTGATCACCACACCCTGCCTCCCAAAGTGAA 3823

Qy   183 GGAATAGTTATATGTACCCCATTTACCCATCACACTCAGTTCACACAGCTTGGTGACATATT 242
         ||||  ||  ||| | |||| ||||| |||  |||||||||  ||||||||||||||||||
Db  3824 AGGATGAAATGATATATGTTCTAATAGCTGTGTCACTCTTTTGTCTTTCAAAAGCAAGTGCAT- 3882

Qy   243 TATTTCTTCTATACCAGTACCGTACCTCCCCACTGGGATTATTTTAAGGCAAAACCCAG 302
         |||| | |||  |  |||| |||| ||||||||||  |||| |||  |||| ||||||||
Db  3883 --TTTGTTAAAAAAAAAAACTTGTCAGGGGCGGGGGCCAGTGGCTTATGCCTGTAATCCC 3940

Qy   303 ATGACATTTTTATCCCTAAATACTTTAGATAAAGGTGTTCTTTGAAAAAAATCATAACCTC 362
         |||||||||||  ||  |||| |||||||  |||| | |||||| ||||||||||||| |||
Db  3941 AATGCTTTGGGAGGCTGAGGCAGGATCACCTGAGGTTGGGAGTTTGA--------- 3990

Qy   363 AGGACAGCAGCCTGCCACACATGTGGAACACCCTGTCTGTACTAAAAAATACAAAAATTAGTT 422
         ||| |  |||||| |||||   || |||  ||||  ||| | |||  |||||||||||||||
Db  3991 -GACCAGCCTGACCAACATGGAGAAACCCCGTCTCTACTAAAAATACAAAAATTAGCCA 4048

Qy   423 GGCCATGGTCGTGGGCGCACCTGTAATCCCAGCTACTCAGGAAGCTGAGGCAGGAGAATCACT 482
         ||| |||| ||||||| ||||| |||||||||||||| | || ||||||||||| ||||| | ||
Db  4049 GACGTGGTGGCGTGCCTGTAGTCCCAGCTACTCGGGAGGCTGAGGCAGGAGAATCACT 4108

Qy   483 TGAATCCGGGAAGCAGAGATTGCAGTGAGCTGAGAT----TGCAGTCGAGCCTGGGCGAC 538
         ||||| ||||| |  |||||  ||||||||| ||||    ||||| | ||||||| |||
Db  4109 GGCACCAAGAGACGAAACTCTGTCCCAAAAAAAAAAAAAAAAAAA 4168
```

```
Query Match         20.1%;    Score 201.4;  DB 10;   Length 7703;
Best Local Similarity  63.1%;  Pred. No. 1.7e-48;
Matches  368;  Conservative   0;  Mismatches 196;  Indels  19;  Gaps  3;

Qy     3 TGCCACCATGCCCTGGCCTAATTTTTCTTAATTTTTTAGTAGAGACGAGGTTTTGCCATGTTGAC 62
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  3644 TGCCACCACGCCCTGGCTAATTTTTGTATTTTTCTAGTAGAGACGGGGTTTCACCATGTTGGC 3703

Qy    63 CAGGCTGGTCTCGAACTCTTGACCTCAGGTGATCCGCCTCAGCCTCCCAAAGTGCT 122
         |||||||||||||||| |||||| |||||||||  ||||||  ||||||||||||||
Db  3704 CAGGCTGGTCTTGAACTCCTGGACCTCAAGTGATCACCACACCCGGCCTCCCAAAGTGCT 3763

Qy   123 GGGATTATAGGCGTGAGCCGCCGCCACCCAGCCAACACATTTTTTAAATACTGGAAAAGTAGAG 182
         ||||||||||||||||||||||  | ||||| |||||||| |||   |||||||||||||||||
Db  3764 GGGATTACAGGCATGAGCCACCGCGCCCGGCCTCAAGTGATCACCACACCCTGCCTCCCAAAGTGAA 3823

Qy   183 GGAATAGTTATATGTACCCCATTTACCCATCACACTCAGTTCACACAGCTTGGTGACATATT 242
         ||||  ||  ||| | |||| ||||| |||  |||||||||  ||||||||||||||||||
Db  3824 AGGATGAAATGATATATGTTCTAATAGCTGTGTCACTCTTTTGTCTTTCAAAAGCAAGTGCAT- 3882

Qy   243 TATTTCTTCTATACCAGTACCGTACCTCCCCACTGGGATTATTTTAAGGCAAAACCCAG 302
         |||| | |||  |  |||| |||| ||||||||||  |||| |||  |||| ||||||||
Db  3883 --TTTGTTAAAAAAAAAAACTTGTCAGGGGCGGGGGCCAGTGGCTTATGCCTGTAATCCC 3940

Qy   303 ATGACATTTTTATCCCTAAATACTTTAGATAAAGGTGTTCTTTGAAAAAAATCATAACCTC 362
         |||||||||||  ||  |||| |||||||  |||| | |||||| ||||||||||||| |||
Db  3941 AATGCTTTGGGAGGCTGAGGCAGGATCACCTGAGGTTGGGAGTTTGA--------- 3990

Qy   363 AGGACAGCAGCCTGCCACACATGTGGAACACCCTGTCTGTACTAAAAAATACAAAAATTAGTT 422
         ||| |  |||||| |||||   || |||  ||||  ||| | |||  |||||||||||||||
Db  3991 -GACCAGCCTGACCAACATGGAGAAACCCCGTCTCTACTAAAAATACAAAAATTAGCCA 4048

Qy   423 GGCCATGGTCGTGGGCGCACCTGTAATCCCAGCTACTCAGGAAGCTGAGGCAGGAGAATCACT 482
         ||| |||| ||||||| ||||| |||||||||||||| | || ||||||||||| ||||| | ||
Db  4049 GACGTGGTGGCGTGCCTGTAGTCCCAGCTACTCGGGAGGCTGAGGCAGGAGAATCACT 4108

Qy   483 TGAATCCGGGAAGCAGAGATTGCAGTGAGCTGAGAT----TGCAGTCGAGCCTGGGCGAC 538
         ||||| ||||| |  |||||  ||||||||| ||||    ||||| | ||||||| |||
Db  4109 TGAACCCAGGAGGTGGCGGTTGCGGTGAGCCGAGATCAAGCCCCATTGCACTCCAGCCTA 4168

Qy   539 AGAGCAGAGAAATGAAAACTCTGTCTCAAAAAACAAAACAAAAACAAAA 581
         ||||||||| |||||||| |||||| |||||||||||| ||||||
Db  4169 GGCAACAAGAGCGAAAACTCTGTCCCAAAAAAAAAAAAAAAAAAAA 4211
```

; Sequence 180, Application US/09954531
; Patent No. US20020165180A1
; GENERAL INFORMATION:
; APPLICANT: Weaver, Zoe
; TITLE OF INVENTION: Process for Identifying Anti-Cancer Therapeutic Agents Using Cand
; TITLE OF INVENTION: Gene Sets
; FILE REFERENCE: 689290-77
; CURRENT APPLICATION NUMBER: US/09/954,531
; CURRENT FILING DATE: 2002-05-02
; PRIOR APPLICATION NUMBER: US/60/233,133
; PRIOR FILING DATE: 2000-09-18
; PRIOR APPLICATION NUMBER: US/60/234,009
; PRIOR FILING DATE: 2000-09-20
; PRIOR APPLICATION NUMBER: US/60/234,034
; PRIOR FILING DATE: 2000-09-20
; PRIOR APPLICATION NUMBER: US/60/234,509
; PRIOR FILING DATE: 2000-09-22
; PRIOR APPLICATION NUMBER: US/60/234,567
; PRIOR FILING DATE: 2000-09-22
; NUMBER OF SEQ ID NOS: 1392
; SOFTWARE: PatentIn version 3.0
US-09-954-531-180
; SEQ ID NO 180
; LENGTH: 65608
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)..(65608)
; OTHER INFORMATION: n=a,t,g or c

Query Match   19.9%;   Score 199.2; DB 9; Length 65608;
Best Local Similarity   63.18;   Pred. No. 2.6e-47;
Matches 381; Conservative   0; Mismatches 203;   Indels 20;   Gaps   4;

Qy   1 CCTGCCACCATGCCTGGCTAATTTCTTATTTTTAGTAGAGACGAGGTTTGCCATGTTG   60
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 35352 CCTGCCACCATGCCTGGCTAATTTCTTATTTTTAGTAGAGACGAGGTTTGCCATGTTG 35411

Qy   61 ACCAGGCTGGTCTCGAACTCTTGACCTCAGGTGATCGCCTGCCTCCCAAAGTG 120
          |||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 35412 GCCAGGCTGATCTTGAACTCCTGACCTCAGGTGATCCCACCCGCCTCGGCCTCCCAAAGTG 35471

Qy   121 CTGGGATTATAGGCGTGAGCCGCGCGCGACCCAGCCA------ACATTTTTTAAATACTGAA 174
          |||||||||||||||||||||| ||| |||| | ||| | |||| |||||||||
Db 35472 CTGGGATTACAGGCGTGAACCACTGTGCCCGGCCATGTACCCGATTATTTTTAACATCATT 35531

Qy   175 AAGTAGAGGGAATAGTT---------ATAGTGTACCCCATTTACCGATTATTTTTAACATCAT 224
          | |||| | ||||| ||||||||||||| |||||||||| |||||
Db 35532 AAGTAGCTGGTATCATTCCCATTTTACAATAAGGAAACTGAGGCTCAGAGAGTCTGTGTC 35591

Qy   225 AACAGCTGGTGACATATTTATTTCTTCTATACCAGTACGTACTCTCCCACTGGGATTA 284
          | ||| |||| || |||| ||||| |||||||| || |||| |||||||
Db 35592 AGTTCCTGAGGTTGCTGTGTAATAAATTGATTATTTTAGAAACAGCAGAAA 35651

Qy   285 TTTTTAAGGCAAAACCCAGATGACATTTTATCCTAAATACTTTAGAT---AAAGGTGTTC 341
          | ||| || |||||||| ||||| |||||| | ||| |||| ||| ||||||||
Db 35652 TGGTCAGGCACAGTGGCTCACACCTGTAATCCCAGCACTTTGGGAGGCCGAGGCGGGCAG 35711

Qy   342 TTTGAAAAAATCATAACCTCAGGACCACCAGCCTGGCCAACACCTGTGAAACCCTGTCTGTAC 401
          ||| |||| ||| | ||| |||| ||| ||| |||| |||||||||||||||||
Db 35712 ATCACTGGAGGTCAGGAGTTCGAGACCAGCCTGGCCAACATGGTGAAACACCATCTCTAC 35771

Qy   402 TAAAAATACAAAAATTAGCTGGGCATGGTGGCACGTGCCTGTAATCCCAGCTACTCAGGA 461
          |||| |||||||||||||||| ||||||||||| |||| ||||||||||| |||||||
Db 35772 TAAAAGTACAAAAATTAGCTGGGCATGGTGGCGCGCGCCTGTAATCCCAGCTACTCGGGA 35831

Qy   462 AGCTGAGGCAGGAGAATCACTTGAATCCGGGAAGCAGCAGAGATTGCAGTGAGCTGAGATTGC 521
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 35832 AATTGAGGCAGGAGAATCGCTTGAACCCAGGAGGCGGAGGTTGCAGTGAGCCGACAATCGT 35891

Qy   522 AGTCGAGCCTGGGCGACAGAGCGAGACAGAAATGAA-ACTCTGTCTCAAAAACAAAACAAA 580
          ||||||||||||||||||||||||||||||| ||| ||| |||||
Db 35892 ACCACTGCACTCTTGCCTGGGACAACAAGAGCAAGACTCCATCTCAAGATAAAAATAAAACAGC 35951

Qy   581 AAAA 584
          | ||
Db 35952 AGAA 35955

RESULT 7
US-09-962-436-292
; Sequence 292, Application US/09962436
; Patent No. US20020081301A1
; GENERAL INFORMATION:
; APPLICANT: Soppet, Daniel
; TITLE OF INVENTION: Cancer Gene Determination and Therapeutic Screening Using Sign
; TITLE OF INVENTION: Sets
; FILE REFERENCE: 689290-75
; CURRENT APPLICATION NUMBER: US/09/962,436
; CURRENT FILING DATE: 2001-09-25
; PRIOR APPLICATION NUMBER: US/60/235,082
; PRIOR FILING DATE: 2000-09-25
; PRIOR APPLICATION NUMBER: US/60/234,924
; PRIOR FILING DATE: 2000-09-25
; NUMBER OF SEQ ID NOS: 568
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 292
; LENGTH: 65608
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
US-09-962-436-292
; OTHER INFORMATION: n=a,t,g or c

Query Match   19.9%;   Score 199.2; DB 10; Length 65608;
Best Local Similarity   63.18;   Pred. No. 2.6e-47;
Matches 381; Conservative   0; Mismatches 203;   Indels 20;   Gaps   4;

Qy   1 CCTGCCACCATGCCTGGCTAATTTCTTATTTTTAGTAGAGACGAGGTTTGCCATGTTG   60
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 35352 CCTGCCACCATGCCTGGCTAATTTCTTATTTTTAGTAGAGACGAGGTTTGCCATGTTG 35411

Qy   61 ACCAGGCTGGTCTCGAACTCTTGACCTCAGGTGATCGCCTGCCTCCCAAAGTG 120
          |||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 35412 GCCAGGCTGATCTTGAACTCCTGACCTCAGGTGATCCCACCCGCCTCGGCCTCCCAAAGTG 35471

Qy   121 CTGGGATTATAGGCGTGAGCCGCGCGCGACCCAGCCA------ACATTTTTTAAATACTGAA 174
          |||||||||||||||||||||| ||| |||| | ||| | |||| |||||||||
Db 35472 CTGGGATTACAGGCGTGAACCACTGTGCCCGGCCATGTACCCGATTATTTTTAACATCATT 35531

Qy   175 AAGTAGAGGGAATAGTT---------ATAGTGTACCCCATTTACCGATTATTTTTAACATCAT 224
          | |||| | ||||| ||||||||||||| |||||||||| |||||
Db 35532 AAGTAGCTGGTATCATTCCCATTTTACAATAAGGAAACTGAGGCTCAGAGAGTCTGTGTC 35591

Qy   225 AACAGCTGGTGACATATTTATTTCTTCTATACCAGTACGTACTCTCCCACTGGGATTA 284
          | ||| |||| || |||| ||||| |||||||| || |||| |||||||
Db 35592 AGTTCCTGAGGTTGCTGTGTAATAAATTGATTATTTTAGAAACAGCAGAAA 35651

Qy   285 TTTTTAAGGCAAAACCCAGATGACATTTTATCCTAAATACTTTAGAT---AAAGGTGTTC 341
          | ||| || |||||||| ||||| |||||| | ||| |||| ||| ||||||||
Db 35652 TGGTCAGGCACAGTGGCTCACACCTGTAATCCCAGCACTTTGGGAGGCCGAGGCGGGCAG 35711

Qy   342 TTTGAAAAAATCATAACCTCAGGACCACCAGCCTGGCCAACACCTGTGAAACCCTGTCTGTAC 401
          ||| |||| ||| | ||| |||| ||| ||| |||| |||||||||||||||||
Db 35712 ATCACTGGAGGTCAGGAGTTCGAGACCAGCCTGGCCAACATGGTGAAACACCATCTCTAC 35771

Qy   402 TAAAAATACAAAAATTAGCTGGGCATGGTGGCACGTGCCTGTAATCCCAGCTACTCAGGA 461
          |||| |||||||||||||||| ||||||||||| |||| ||||||||||| |||||||
Db 35772 TAAAAGTACAAAAATTAGCTGGGCATGGTGGCGCGCGCCTGTAATCCCAGCTACTCGGGA 35831

Qy   462 AGCTGAGGCAGGAGAATCACTTGAATCCGGGAAGCAGCAGAGATTGCAGTGAGCTGAGATTGC 521
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 35832 AATTGAGGCAGGAGAATCGCTTGAACCCAGGAGGCGGAGGTTGCAGTGAGCCGACAATCGT 35891

Qy 522 AGTCGAGCCTGGGCGGACAGAGACGAGAAATGAA-ACTCTCTCTCAAAAAACAAACAAA 580
         || |||||||||||||||||||||||||||||||| || |||||||| ||||| |||||
Db 35892 ACCACTGCACTCTTGCCTGGACAACAAAGCAAGACTCCATCTCAAGATAAAAATAAACAGC 35951

Qy 581 AAAA 584
         | ||
Db 35952 AGAA 35955

RESULT 8
US-09-962-832-119
; Sequence 119, Application US/09962832
; Patent No. US20020110821A1
; GENERAL INFORMATION:
; APPLICANT: Ebner, Reinhard
; TITLE OF INVENTION: Cancer Gene Determination and Therapeutic Screening Using Signatu
; TITLE OF INVENTION: Sets
; FILE REFERENCE: 68929O-74
; CURRENT APPLICATION NUMBER: US/09/962,832
; CURRENT FILING DATE: 2001-09-25
; PRIOR APPLICATION NUMBER: US/60/235,077
; PRIOR FILING DATE: 2000-09-25
; PRIOR APPLICATION NUMBER: US/60/235,280
; PRIOR FILING DATE: 2000-09-25
; NUMBER OF SEQ ID NOS: 259
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 119
; LENGTH: 65608
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; OTHER INFORMATION: n=a,t,g or c
US-09-962-832-119

Query Match 19.9%; Score 199.2; DB 10; Length 65608;
Best Local Similarity 63.1%; Pred. No. 2.6e-47;
Matches 381; Conservative 0; Mismatches 203; Indels 20; Gaps 4;

Qy   1 CCTGCCACCATGCCTGGCTAATTTCTTTATTTTTAGTTAGTTAGAGACGAGAGGTTTGCCATGTTG 60
Db 35352 CCTCCACCATGCCTGGCTAATTTTTTCTTTATTTTTAGTTAGTACAGATGGGGTTTCACCATGTTG 35411

Qy   61 ACCAGGCTGGTCTCGAACTCTTGACCTCAGGTGATCGGCCGCTCCAGCCTCCCAAAGTG 120
Db 35412 GCCAGGCTGGTCTTGAACTCCTCAGACCTCAGGTGATCACCGCGCCTCCCAAAGTG 35471

Qy 121 CTGGGATTATAGGCGTGAGCCGCGCCACCCAGCCA------ACATTTTTTAAATACTGAA 174
Db 35472 CTGGGATTACAGGTGTGAACCACTGTGCCCGGCCGATTATTTTTAACATCATT 35531

Qy 175 AAGTAGAGGGAATAGTT--------ATAGTGTACCCCATTTACCCATCACTCACTCAGTTTC 224
Db 35532 AAGTAGCTGTGTATCCATTTTACTAATAAGGAGAAACTGAGGGCTCAGAGAGTCGTGTGTG 35591

Qy 225 AACAGCTGGTGACATATTTATTTCTTCTATACCAGTACGTACTCTCCCACTGGGATTA 284
Db 35592 AGTTTCCTGAGGTTGCTGTGTAATAAAATGTTAGGAAAACTTGATTATTTAAAACAGCAGAAAA 35651

Qy 285 TTTTAAGGCAAAACCCAGATGACATTTTATCCCTAAATACTTTAGAT----AAAGGTGTTC 341
Db 35652 TGGTCAGGCGACAGTGGCTCACACCTGTAATCCCAGCACTTTGGGAGGCCGAGGGGGCAG 35711

Qy 342 TTTGAAAAAATCATAACCCTCAGGACCAGCAGCCTCGGCCAACATGTGAAACCCGTCTGTAC 401
Db 35712 ATCACTGGAGGTCAGGAGTTCGAGACCAGCCTGGCCAACATGGTGAAACCACCATCTCTAC 35771

Qy 402 TAAAAAATACAAAAATTAGCTTGGCATGGTGGCGTGGGCGCGTGGGCCACCTGTAATCCCAG 461
Db 35772 TAAAAGTACAAAAATTAGCTGGGCGCGGTGGCCAGGGGCCTGTAATCCCAGCTACTCGGGA 35831

Qy 462 AGCTGAGGCAGGAGAATCACTTGAATCCGGGAGGCAGAGATTGCAGTGAGCTGAGATTGC 521
Db 35832 CTTGAACCCGGGAGGCGGAGGTTGCAGTGAGCCGAGATCATGCCACTGCACTCCAGCCTG 35891

Db 35832 AATTGAGGCAGGAGAATCGCTTGAACCCAGGAGGCGGAGGTTGCAGTGAGCCGAGATCGT 35891

Qy 522 AGTCGAGCCTGGGCGGACAGAGACGAGAAATGAA-ACTCTGTCTCAAAAAACAAACAAA 580
Db 35892 ACCACTGCACTCTTGCCTGGACAACAAAGCAAGACTCCATCTCAAGATAAAAATAAACAGC 35951

Qy 581 AAAA 584
Db 35952 AGAA 35955

RESULT 9
US-10-000-639-7/c
; Sequence 7, Application US/10000639
; Publication No. US20020192777A1
; GENERAL INFORMATION:
; APPLICANT: SHEPPARD, PAUL O.
; APPLICANT: VU, TUYEN Q.
; APPLICANT: FELDHAUS, ANDREW L.
; APPLICANT: HALDEMAN, BETTY A.
; TITLE OF INVENTION: Testis Protein, Zsig986
; FILE REFERENCE: 00-44
; CURRENT APPLICATION NUMBER: US/10/000,639
; CURRENT FILING DATE: 2001-11-01
; PRIOR APPLICATION NUMBER: 60/245,070
; PRIOR FILING DATE: 2000-11-01
; NUMBER OF SEQ ID NOS: 16
; SEQ ID NO 7
; LENGTH: 9887
; TYPE: DNA
; SOFTWARE: FastSEQ for Windows Version 3.0
; ORGANISM: Homo sapiens
US-10-000-639-7

Query Match 19.5%; Score 194.8; DB 9; Length 9887;
Best Local Similarity 63.8%; Pred. No. 1.7e-46;
Matches 369; Conservative 0; Mismatches 187; Indels 22; Gaps 4;

Qy   1 CCTGCCACCATGCCTGGCTAATTTCTTTATTTTTAGTTAGTTAGAGACGAGAGGTTTGCCATGTTG 60
Db 7458 CCTGCTCCATGCCTGGCTAATTTTTTCTTTATTTTTAGTTAGTACAGATGGGGTTTCACCATGTTG 7399

Qy   61 ACCAGGCTGGTCTCGAACTCTTGACCTCAGGTGATCGGCCGCTCCAGCCTCCCAAAGTG 120
Db 7398 ACCAGGCTGGTCTCGAACTCCTGACCTCAGGTGATCCGCCCGCCTCGGCCTCCCAAAGTC 7339

Qy 121 CTGGGATTATAGGCGTGAGCCGCGCCACCCAGCCA------ACATTTTTTAAATACTGAA 180
Db 7338 CTGGGATTACAGGCGTGAGCCACCGCGCCC-----TGTTCTATTCATTATTGAAAG 7285

Qy 181 AAGGGAATAGTTATAGTGTACCCCATTTACCCATCACTCAGTTTCAACAGCTGGTGACATA 240
Db 7284 TGGGGGGTTGAAATAATGAAACTAATAAATTGTAGAGCTGACTACTTTCAATTCTGATATTTT 7225

Qy 241 TTTATTTCTTCTATACCAGTACGTACTCTCCCCACTGGGATTATTTATTTAAGGCAAAACCC 300
Db 7224 TGTTTCATTTATTTTTGGGGTCTCCGTTAGGCACATAAAATGTTTCTAATTGTCATATATTGCCT 7165

Qy 301 AGATGACATTTTATCCCTAAATACTTTAGATAAAGGTGTTCTTTGAAAAAATCATTAACC 360
Db 7164 TGTT----------TTGAAAACTTTTATTAGGTGGGGATCACCTCAGGTGCAGGAGT 7118

Qy 361 TCAGGACCAGCCTGGCCAACATGGTGAAACCCGTCTGTACTACAAAAAATACAAAAATTAGC 420
Db 7117 TCAAGACCAGCCTGACCAACATGGATAAAACCCCGTCTCTACTAAAAATAC-AAAAATTAGT 7059

Qy 421 TTGGCATGGTGGCGTGGGCGCCTGTAATCCCAGCTACTCGGGAGGCTGAGGCAGGAGAATCA 480
Db 7058 AGGGCATGGTGGCGCATGCCTGTAATCCCAGCTACTCAGGAGGCTGAGGCAGGAGAATTG 6999

Qy 481 CTTGAATCCGGGAGGCAGAGATTGCAGTGAGCTGAGAT--TGCAGTCGAGCCTGGGCGAC 538
Db 6998 CTTGAACCCGGGAGGCGGAGGCTGTGGTGAGCCAAGATCATGCCACTGCACTCCAGCCTG 6939

```
Qy    539  AGAGACAGAAAATGAAAACTCTGTCTCAAAAAACAAACAAA  576
           |||  |||||||||||||||||||||||||||||||| |||
Db   6938  GCCAACAAGAGCGAAACTCTGTCTCAAAAAAAAAAAAA    6901

RESULT 10
US-09-801-574-61
; Sequence 61, Application US/09801574
; Patent No. US20020815924A1
; GENERAL INFORMATION:
;   APPLICANT: Wang, Peijing Jeremy
;   APPLICANT: Page, David C.
;   TITLE OF INVENTION: Reproduction-Specific Genes
;   FILE REFERENCE: 0399.2007-002
;   CURRENT APPLICATION NUMBER: US/09/801,574
;   CURRENT FILING DATE: 2001-03-07
;   PRIOR APPLICATION NUMBER: 60/187,518
;   PRIOR FILING DATE: 2000-03-07
;   PRIOR APPLICATION NUMBER: 60/261,557
;   PRIOR FILING DATE: 2001-01-12
;   NUMBER OF SEQ ID NOS: 90
;   SOFTWARE: FastSEQ for Windows Version 4.0
;   SEQ ID NO 61
;   LENGTH: 53542
;   TYPE: DNA
;   ORGANISM: Homo sapiens
US-09-801-574-61

Query Match          19.1%;   Score 191.2;  DB 10;  Length 53542;
Best Local Similarity  60.6%;   Pred. No. 5.2e-45;
Matches    383;  Conservative   0;  Mismatches  218;  Indels  31;  Gaps  3;

Qy      5  CCACCATGCCTGGCTAATTTCTTATTTTTAGTAGACGAGGTTTGCCATGTTGACCA   64
           |||||||||||||||||||||||| ||| ||| || | ||||| ||| |||| |||
Db  26067  CCATCATGCCCGGCTAATTTTTGGATTTTTTAGTAGAGATGGGGTTTCACCATGTTGGCCA  26126

Qy     65  GGCTGGTGTCTCGAACTCTTGACCTCAGGTGATCCGCCTCCAGCCTCCCAAAGTGCTGG  124
           ||||||| || || ||| |  |||| |||||| ||| | |||| ||||||| || || |
Db  26127  GGCTGGTCTTGAACTCCTGATCTCA--TGATCTGCCCGCCTCCGGCCTCCCAAAGTGCTGG  26184

Qy    125  GATTATAGGCGTGAGCCGCGCGCCACCCAGCCCACACATTTTTTAAATA------CTGAAAG  177
           ||||||||| || |||||| |||| ||||||||||| ||||  |||| |      |||||||
Db  26185  GATTACAGGCGTGAGCCACTGCCACCCAGGCCTAGAAGTTTTTATAAAATATCTTTGCAACA  26244

Qy    178  TAGAGGGGAATAGTTATAGTGTACCCCATTTACCCAGTTTCAACAGCTGGTGAC  237
           | ||| | ||  || | ||| || |||| | |||| | ||| || |||| |||
Db  26245  TTCACAAAGTTTGCAAATGTTAACAATTGCGGGGGTGTATCCTTCCAGACATTTTCCA  26304

Qy    238  ATATTTATTTCTTCTATACCAGTACGTACTCTCCCACTGGGATTATTTTTAAGGCAAAA  297
           | ||| | ||  || | || | | || || ||| | ||| |||  | |||| | || |
Db  26305  TGCATACACAATCACATAAAATGTACCTAAGTGTACACCATATAAACCATATAAATTAAAAATA  26364

Qy    298  ------------CCCAGAATGACATTTTTATCCTTAAATACTTTAGATAAAG  335
                       || | ||| || || | |||| |||  ||| | ||||
Db  26365  AAACTCTAAGGCGCGGGTGTTGGTGGCCTCAGGCCTCGTAATTCCAGCACTTTGGGGAGGCCAG  26424

Qy    336  GTGTTCTTTGAAAAAATCATAACCTCAGGACCAGCCAGCCAGCCTGT  395
           ||   || |||| | | | ||||| ||| || ||| |||||||| |
Db  26425  GAGGGTGGATCACAAGATCAGGAGAGATCGAGACCAGCCTGACCAATATGGTGAAACCCTGT  26484

Qy    396  CTGTACTAAAAAATACAAAAAATTAGCTTGGCATGGTGGTCGTGGGCACCTGTAATCCCAGCTAC  455
           ||| ||||||||||||||||||| ||| || ||| ||||  | | ||||| ||| |||| | |||
Db  26485  TTCTACTAAAAAATACAAAAAATTAGCTGGGCGTGGTGGCACATGCCTGTAGTCCTAGCTAC  26544

Qy    456  TCAGGGAAGCTGAGGCAGGAGAATCACTTGAATCCGGGAAGGCAGGAGGTTGCAGTGAGCTGA  515
           || || ||||||||| || ||| ||| |  | ||||||| |||| | | ||  | || |||
Db  26545  TCTGGAGACTGAGACAGGAGAATCACTTGAATCTGGGAGGCAGGAGGTTGCAGTGAGCCAA  26604

Qy    516  GATTGCAGTGCGAGCCTGGGCGACAGAGCGAGAAATGAAAACTCTGTCTCAAAAAACAAACAA  575
           ||||||||  | |||| ||||||| ||| | | |||| |||||||||||||||||||||||| |
Db  26605  GATTGCACCACTGCACTCCAGCCTGGGTGAAGAGCGAGACTCTGTCTCAAAAAAAAAAAGA  26664
```

```
Qy    576  ACAAAAAAACCACTATACATAAAAAATGAACAA  607
           |  |||| ||  |  |  |  |   | |||| |||
Db  26665  AAGAAAAGAGAAAAGAAAAAGAAAAAAAAACCCAA  26696

RESULT 11
US-09-933-267A-1/c
; Sequence 1, Application US/09933267A
; Patent No. US20020123095A1
; GENERAL INFORMATION:
;   APPLICANT: Kalush, Francis et al.
;   TITLE OF INVENTION: Estrogen receptor alpha variants and
;   TITLE OF INVENTION: methods of detection thereof
;   FILE REFERENCE: CL00025BC14
;   CURRENT APPLICATION NUMBER: US/09/933,267A
;   CURRENT FILING DATE: 2001-08-21
;   PRIOR APPLICATION NUMBER: 60/160626
;   PRIOR FILING DATE: 1999-10-20
;   PRIOR APPLICATION NUMBER: 60/183756
;   PRIOR FILING DATE: 2000-02-22
;   PRIOR APPLICATION NUMBER: 09/692414
;   PRIOR FILING DATE: 2000-10-20
;   PRIOR APPLICATION NUMBER: 09/768184
;   PRIOR FILING DATE: 2001-01-24
;   PRIOR APPLICATION NUMBER: 09/804076
;   PRIOR FILING DATE: 2001-03-13
;   PRIOR APPLICATION NUMBER: 09/826314
;   PRIOR FILING DATE: 2001-04-05
;   NUMBER OF SEQ ID NOS: 2
;   SOFTWARE: FastSEQ for Windows Version 4.0
;   SEQ ID NO 1
;   LENGTH: 465237
;   TYPE: DNA
;   ORGANISM: human
US-09-933-267A-1

Query Match          19.0%;   Score 189.6;  DB 10;  Length 465237;
Best Local Similarity  61.6%;   Pred. No. 5.4e-44;
Matches    373;  Conservative   0;  Mismatches  199;  Indels  34;  Gaps  3;

Qy      5  CCACCATGCCTGGCTAATTTCTTATTTTTTAGTAGACGAGGTTTTGCCATGTTGACCA   64
           ||||||||| || ||||||||| | | ||| || | ||||  ||| | |||| |||
Db 119149  CCACCACACCCAGCTAATTTTTGTATTTTTTAATAGAGACGGGGTTTCACCATGTTGCCCA  119090

Qy     65  GGCTGGTCTCGAACTCTTGACCTCAGGTGATCCGCCTCCAGCCTCCCAAAGTGCTGG  124
           ||||||| || || ||| | ||| |  ||  | | | | || ||||| ||||||| |
Db 119089  GGCTGGTCTTGAACTCCTGACCTCAAGTGATCTGCCCGCCTCCAAAGTGCTGG  119030

Qy    125  GATTATTAGGCGTGAGCCGCGCGCCACCCAGCCCAACACATTTTTTAAATACTGAA--------  174
           ||||| ||| || |||||| |||| ||||||||| | ||  |||||||| |
Db 119029  GATTACAGGCGTGAGCCATCACGCCCGGCCCATTTAATTTTATTTTCACCAGTTAGAATTT  118970

Qy    175  --------AAGTAGAGGGAATAGTTGTACCCAGCAATGTTTTTAGTGTACCCATCACTCAGTTTCAA  226
                  ||||| | |||| | ||| ||| ||| | | || | | |||| | ||
Db 118969  GTTACTCTCAATTTATTCAAAAAACTATAATTGACAGTACTACCTACGGTTATA  118910

Qy    227  CAGCTGGGTGACATATTTATTTCTCTATACCAGTACGTTACCTCCCAATGGGGATTATT  286
           ||||||| ||| || | | | | ||  |||| | | ||||||  | | ||||| | |
Db 118909  AATATTGGAAATAGAGCAATTGGACAAAACAGAAGTCTACCCTCATACCAGCTTATATT  118850

Qy    287  TTAAGGCA---------------AAACCCAGATGACATTTTTATCCCTAAATACTTTTAGAT  331
           || |||||               |||||| | || ||||| |||| ||| | ||| |||
Db 118849  CTAGTGTAGAGACAGACAGACAAATAATTAGGTAGCATACACAGGAGGTTAGAAGTAATAC  118790

Qy    332  AAAGGTGTTGTTTTGAAAAAAAAATCATAACCTCAGACAACCTTGGCTGGGTGAAACC  391
           |||| || || ||  ||||||||| ||||| |||| ||   |||| | ||||| ||
Db 118789  CAGCACTTTGGGAGGCCAGGTCAGGAGTTGGGAGACCAGCCTGATGGAACC  118730

Qy    392  CTGTCTGTACTAAAAATACAAAAAATTTAGCTGGCATGGTGGTCGTGGGCACCTGTAATCCCAG  451
           ||| | ||| ||||||| |||||| ||| || |||| |||  | |  |||| ||| |||||||
Db 118729  CTGTCTCTACTAAAAATACAAAAAATTAGCCAGGCATGGTGGCACGCGCCTGTAATCCCAG  118670

Qy    452  CTACTCAGGAAGCTGAGGCAGGAGAATCACTTGAATCCGGGAGGAAGCGAGAGATTGCAGTGAG  511
```

RESULT 12

```
Db 118669 CTACTGCGGGAGGCCGAGCGGCGGGCGACAGAGACAGA-AATGAAACTCTGTCTCAAAAACA 118610
Qy    512 CTGAGATTGCCAGTCGAGCGCCGAGCCGGACAGAGACAGA-AATGAAACTCTGTCTCAAAAACA 570

Db 118609 CCAAGAATTGCGCCACTGCGCTCCAGCCTAAGTGACAGAGTAATACTCTGTCTCAAAAAGA 118550
Qy    571 AACAAA 576

Db 118549 AAAGAA 118544
```

US-10-016-725-15/c
Sequence 15, Application US/10016725
Patent No. Us20020151018A1
; GENERAL INFORMATION:
;   APPLICANT: Prouty, Stephen
;   APPLICANT: Zhang, Lin
;   APPLICANT: Stenn, Kurt
;   TITLE OF INVENTION: Stearyl-CoA Desaturase Promoter
;   FILE REFERENCE: J&J20065
;   CURRENT APPLICATION NUMBER: US/10/016,725
;   CURRENT FILING DATE: 2002-06-10
;   NUMBER OF SEQ ID NOS: 22
;   SOFTWARE: Patentin version 3.1
;   SEQ ID NO 15
;   LENGTH: 4150
;   TYPE: DNA
;   ORGANISM: Homo sapiens
US-10-016-725-15

Query Match   18.9%; Score 189.4; DB 12; Length 4150;
Best Local Similarity   64.0%;
Matches 372; Conservative 0; Mismatches 191; Indels 18; Gaps 5;

```
Qy     4 GCCACCATGCCTGGCTAATTTTCT-TATTTTTAGTGAGAGACGAGGTTTTGCCATGTTGAC 62
Db  2008 GCCACCACACTTAGCTAATTTTTTGTATTTTTAGTAGAGACGGGGTTTCACCATGTTGGC 1949

Qy    63 CAGGCTGGTCTCGAACTCCTGACCTCAGGTGATCCGCCTGCCTCAGCCTCCCAAAGTGCT 122
Db  1948 CAGGCTGGTTTGAACTCCTGACCTCAGGTGATCCACATGCCTGGCCTCCCAAAGTGCT 1889

Qy   123 GGGATTATAGGCGTGAGCCGCCGCGCCCACCCAGCC---AACATTTTTTAAATACTGAAAAGTA 179
Db  1888 GGGATTACAGGCATGAGCCCACCACGCCTGGCCCGGAGAATATTTTCATTTCTATTTCTT 1829

Qy   180 GAGGGAATAGTTATAGTGTACCCCATTTACCCATCACTCAGTTTCAACAGCTGGTGACAT 239
Db  1828 TCGAGCAGATCACCTCTACTCCTGAGGCCCTGGTCACACTGGTTGTTCTCAAGTGACAG 1769

Qy   240 ATTTATTTCTTCTATACCAGTACCGTACTCCTCCCCACTGGGATTATAGTTATTTTTAAGGCAAAAACC 299
Db  1768 TGGT------CAAACCAACACACTGATTTGGATTCTTAAGAGATTTATTTTTTTGGTCTG 1717

Qy   300 CAGATGACATTTTATCCCTAAATACTTTAGATAAAGGTGTTCTTTGAAAAAAATCATAAC 359
Db  1716 GGGGCCCTACCTCAGCAAAAGGAACAATAGTTACAACAATTATGACAGCAAA-----AAG 1662

Qy   360 CTCAGGACCAGCGCTGGCCAACATGGTGAAAACCCTGTCTGTACTAAAAATACAAAAATTAG 419
Db  1661 TTTGAGACCAGCTTGGCCAACAAAGTGAAAATCCCATCTCTACTAAAAATACAAAAATCAG 1602

Qy   420 CTTGGCCATGGTGGTGCGTGGGCCACCTGTAATCCCAGCTACTCAGGAAGCTGAGGCAGGAGAATC 479
Db  1601 CTGGGCATGGTGGGGAAGTGCCCTGTAGTCCCAGCTACTTGGGAGGCTGAGGCAGGAGAATT 1542

Qy   480 ACTTGAATCCGGGAGGCAGAGATTGCAGTGAGCTGAGATTGCAGTCGACCCTGGCGACA 539
Db  1541 GCTTGAACCCGGGAGGCGGAAAGTTGCAGTGAGCTGAGATCACACCACTGCCCTCCAGCCA 1482

Qy   540 GAGACAGAAAATGAA-ACTCTGTCTCAAAAAACAAACAAACAA 579
```

RESULT 13

```
Db  1481 GGGTGAAAGACCAAGACTCTTTTCTCAAAAAAAAAAAAAAAA 1441
```

US-09-954-531-180/c
Sequence 180, Application US/09954531
Patent No. US20020165180A1
; GENERAL INFORMATION:
;   APPLICANT: Weaver, Zoe
;   TITLE OF INVENTION: Process for Identifying Anti-Cancer Therapeutic Agents Using c
;   FILE REFERENCE: 689290-77
;   CURRENT APPLICATION NUMBER: US/09/954,531
;   CURRENT FILING DATE: 2002-05-02
;   PRIOR APPLICATION NUMBER: US/09/954,531
;   PRIOR FILING DATE: 2000-09-18
;   PRIOR APPLICATION NUMBER: US/60/233,133
;   PRIOR FILING DATE: 2000-09-18
;   PRIOR APPLICATION NUMBER: US/60/234,009
;   PRIOR FILING DATE: 2000-09-20
;   PRIOR APPLICATION NUMBER: US/60/234,034
;   PRIOR FILING DATE: 2000-09-22
;   PRIOR APPLICATION NUMBER: US/60/234,509
;   PRIOR FILING DATE: 2000-09-22
;   PRIOR APPLICATION NUMBER: US/60/234,567
;   PRIOR FILING DATE: 2000-09-22
;   NUMBER OF SEQ ID NOS: 1392
;   SOFTWARE: Patentin version 3.0
;   SEQ ID NO 180
;   LENGTH: 65608
;   TYPE: DNA
;   ORGANISM: Homo sapiens
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (1)...(65608)
;   OTHER INFORMATION: n=a,t,g or c
US-09-954-531-180

Query Match   18.9%; Score 189; DB 9; Length 65608;
Best Local Similarity   60.1%;
Matches 377; Conservative 0; Mismatches 230; Indels 20; Gaps 3;

```
Qy     5 CCACCATGCCTGGCTAATTTTCTTTATTTTTAGTAGAGACGAGGTTTTGCCATGTTGACCA 64
Db 27147 CCACCAAAACCCAGCTAATTTTTGTAGTTTTAGTAGAGACGGGGTTTCACCATGTTGCCA 27088

Qy    65 GGCTGGTCTCGAAACTCCTTGACCTCAGGTGATCCGCCTGCCTCCGCCTCCCAAAGTGCTGG 124
Db 27087 GGCTGTTCTCAAAACTCCTACCTCAAGTGAGTCCACCTCGCCTCAGCCTCCCAAAGTGCTGG 27028

Qy   125 GATTATAGGCGTGAGCCGCCGCGCCCAGCC------AACATTTTTTAAATACTGAAAAGT 178
Db 27027 CATTACAGGCGTGAGCCACCGCACCCCGGCCCCCTACAAAACATTATTTCCTAAGACACTGA 26968

Qy   179 AGAGGGAATAGTTATAGTGTACCCCATTTACCCATCACTCAGTTTCAACAGCTGGTGACA 238
Db 26967 TTTCAAACTCTAAGACAGGGATAATTCCTCTATAGCTCCCAAATCTGCCTCCTGAGAAAAC 26908

Qy   239 TATTTATTTCTTCTATACCAGTACCGTACCTCCCCACTGGGATTATTTGAAGGCAAAAC 298
Db 26907 AAGCTATATTTATTTATTTCTATAAAATCTCATCT-----TTCTGTCAGGTGTGTG 26859

Qy   299 CCAGATGACATTTTATCCCTAAATACTTTAGATAAAGGTGTTCTTTGAAAAAAATCATAA 358
Db 26858 GCTCACGCCTGTAATCCCAACACTTTGGGAGGTCAGGAGTTCACGAGGTCAGGA 26799

Qy   359 CCTCAGGACCAGCCTGCCAACATGGTGAAACCCCGTCTGTACTAAAAATACAAAAATTA 418
Db 26798 GTTTGAGACCAGCCTGACCAACATGGTGAAACCCCGTCTCTACTAAAAATACAAAAATTA 26739

Qy   419 GCTTGGCATGGTGGCGGGCGCACCTGTAATCCCAGCTACTCAGGAAGCTGAGGCAGGAGAAT 478
Db 26738 GCCGGCGTGTGGTGGCGGGCGCCTGTAATCCCAGCTACTCGGGAGGCTGAGGC---AGAAT 26682
```

```
Qy    479 CACTTGAATCCGGGAAGCAGAGAATTGCAGTGAGTCGAGCCTGGGGCGGAC 538
          |||| ||  ||||||| |  |   ||||| |||||||| || | ||| |
Db  26681 CGATTGAACCTGGGAGGTGGAGTTGCAGTGAGTCACGCGAGATTGCACTTCAGC 26622

Qy    539 AGAGACAGAGAATGAAACTGCTGTCTCAAAAAACAAACAAACAACTATACATAA 598
          |||| ||  || ||||| ||||| || |||||||| | ||||||| | |||||
Db  26621 TGGACAGTAAGAGAGCAAAAACTCCATCTCGGGAAAAACAAAACAAAAAACCACTATACATAA 26562

Qy    599 AATGAACAAATGATGACGCCACAATACCACC 625
          ||| ||  ||  |||  ||  | |||| |||
Db  26561 TTTCTAGGTTTTTAAAAAAATTAGAACC 26535
```

RESULT 14
US-09-962-436-292/c
Sequence 292, Application US/09962436
Patent No. US20020081301A1
GENERAL INFORMATION:
; APPLICANT: Soppet, Daniel
; TITLE OF INVENTION: Cancer Gene Determination and Therapeutic Screening Using Signatu
; TITLE OF INVENTION: Sets
; FILE REFERENCE: 689290-75
; CURRENT APPLICATION NUMBER: US/09/962,436
; CURRENT FILING DATE: 2001-09-25
; PRIOR APPLICATION NUMBER: US/60/235,082
; PRIOR FILING DATE: 2000-09-25
; PRIOR APPLICATION NUMBER: US/60/234,924
; PRIOR FILING DATE: 2000-09-25
; NUMBER OF SEQ ID NOS: 568
; SOFTWARE: Patentin version 3.0
; SEQ ID NO 292
; LENGTH: 65608
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; OTHER INFORMATION: n=a,t,g or c
US-09-962-436-292

Query Match   18.9%; Score 189; DB 10; Length 65608;
Best Local Similarity 60.1%; Pred. No. 2.6e-44;
Matches 377; Conservative 0; Mismatches 230; Indels 20; Gaps 3;

```
Qy      5 CCACCATGCCTGGCTAATTTTTGTATTTTTAGTTAGAGACGAGGTTTGGCCATGTTGACCA 64
          |||||   |  ||||||||||||||||||||||||||||||||||| ||||| ||| ||
Db  27147 CCACCAAACCCAGCTAATTTTTGTATTTTTAGTTAGAGACGAGGTTTCACCATGTTGGCCA 27088

Qy     65 GGCTGGTCTCGAACTCTTGACCTCAGGTGATCCGCCTGCCTCCCAAAGTGCTGG 124
          ||||| ||||| ||||| | ||||| ||||| |  ||||||  ||||||||||
Db  27087 GGCTGTTCTCAAACTCCTAACCTCAAGTGATCACCTGCCTCGGCCTCCCAAAGTGCTGG 27028

Qy    125 GATTATAGGCGTGAGCCGCCGCGACCCAGCC-----AACATTTTTAAATACTGAAAAGT 178
          ||||||||| ||||||||||| |||||||||     |||||||||||||||| ||||
Db  27027 CATTACAGGCGTGAGCCGACCCGGACCCGACCCAGCCAACATTATTTCCTAACAGAACTGA 26968

Qy    179 AGAGGGAATAGTTATATAGTGTACCCCCATTTACCCATCACTCAGTTTCAACAGCTGGTGACA 238
          ||||| |||||||||||||||| ||||||||||||||||| ||||  || || |||| |||
Db  26967 TTTCAAACTCTAAGCAGGGATAATCCTCTATAGCTCCCAAATCTGCCTCTCGAGAAAC 26908

Qy    239 TATTTATTTCTTCTATACCAGTACCGTACTCTCCCACTGGGATTATTTTAAAATCTCATCT---- 298
          ||||||||||| |||||||||| |||||| ||| ||||||||||| ||||||||||
Db  26907 AAGCTATATTTATTTTTCTATAAAATCTCATCT----TTCTGTCAGGTGTGGTG 26859

Qy    299 CCAGATGACATTTTATCCCTAAATACTTTAGATAAAAGGTGTT 358
          ||||| ||||||| ||||| |||| ||| | |||| |||||||
Db  26858 GCTCACGCCTGTAATCCCAACACTTTGGGAGGTCAGGA 26799

Qy    359 CCTCAGGACCAGCCTGGCCAACACATGGTGAAAACCCTGTCT 418
          |||||| |||||||||| |||||| |||||||||||||||||
Db  26798 GTTTGAGACCAGCCTGACCAACATGGTGAAAACGCAAAAAATTA 26739

Qy    419 GCTTGGCATGGTGGTGCGTGGGCCACCTGTAATCCCAGCTACTCAGGAGGCTGAGGCAGGAGAAT 478
          |||||||| |||| ||| |||| |||||||||||
Db  26798 GTTTGAGACCAGCCTGACCAACATGGTGAAAACGCAAAAAATTA 26739
```

```
Db  26738 GCCGGGGTGTGGTGGTGGGGCTCCTGTAATCCCAGCTACTCGGGAGGCTGAGGC---AGAAT 26682

Qy    479 CACTTGAATCCGGGAAGCAGAGAATTGCAGTGAGTCGAGCCTGGGGCGGAC 538
          |||| ||  ||||||| |  |   ||||| |||||||| || | ||| |
Db  26681 CGATTGAACCTGGGAGGTGGAGTTGCAGTGAGTCACGCGAGATTGCACTTCAGC 26622

Qy    539 AGAGACAGAGAATGAAACTGCTGTCTCAAAAAACAAACAAACAACTATACATAA 598
          |||| ||  || ||||| ||||| || |||||||| | ||||||| | |||||
Db  26621 TGGACAGTAAGAGAGCAAAAACTCCATCTCGGGAAAAACAAAACAAAAAACCACTATACATAA 26562

Qy    599 AATGAACAAATGATGCCACAATAGCACC 625
          ||| ||  ||  |||  ||  | |||| |||
Db  26561 TTTCTAGGTTTTTAAAAAAATTAGAACC 26535
```

RESULT 15
US-09-962-832-119/c
Sequence 119, Application US/09962832
Patent No. US20020110821A1
GENERAL INFORMATION:
; APPLICANT: Ebner, Reinhard
; TITLE OF INVENTION: Cancer Gene Determination and Therapeutic Screening Using Sign
; TITLE OF INVENTION: Sets
; FILE REFERENCE: 689290-74
; CURRENT APPLICATION NUMBER: US/09/962,832
; CURRENT FILING DATE: 2001-09-25
; PRIOR APPLICATION NUMBER: US/60/235,077
; PRIOR FILING DATE: 2000-09-25
; PRIOR APPLICATION NUMBER: US/60/235,280
; PRIOR FILING DATE: 2000-09-25
; NUMBER OF SEQ ID NOS: 259
; SOFTWARE: Patentin version 3.0
; SEQ ID NO 119
; LENGTH: 65608
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; OTHER INFORMATION: n=a,t,g or c
US-09-962-832-119

Query Match   18.9%; Score 189; DB 10; Length 65608;
Best Local Similarity 60.1%; Pred. No. 2.6e-44;
Matches 377; Conservative 0; Mismatches 230; Indels 20; Gaps 3;

```
Qy      5 CCACCATGCCTGGCTAATTTTTGTATTTTTAGTTAGAGACGAGGTTTGGCCATGTTGACCA 64
          |||||   |  ||||||||||||||||||||||||||||||||||| ||||| ||| ||
Db  27147 CCACCAAACCCAGCTAATTTTTGTATTTTTAGTTAGAGACGAGGTTTCACCATGTTGGCCA 27088

Qy     65 GGCTGGTCTCGAACTCTTGACCTCAGGTGATCCGCCTGCCTCCCAAAGTGCTGG 124
          ||||| ||||| ||||| | ||||| ||||| |  ||||||  ||||||||||
Db  27087 GGCTGTTCTCAAACTCCTAACCTCAAGTGATCACCTGCCTCGGCCTCCCAAAGTGCTGG 27028

Qy    125 GATTATAGGCGTGAGCCGCCGCGACCCAGCC-----AACATTTTTAAATACTGAAAAGT 178
          ||||||||| ||||||||||| |||||||||     |||||||||||||||| ||||
Db  27027 CATTACAGGCGTGAGCCGACCCGGACCCGACCCAGCCAACATTATTTCCTAACAGAACTGA 26968

Qy    179 AGAGGGAATAGTTATATAGTGTACCCCCATTTACCCATCACTCAGTTTCAACAGCTGGTGACA 238
          ||||| |||||||||||||||| ||||||||||||||||| ||||  || || |||| |||
Db  26967 TTTCAAACTCTAAGCAGGGATAATCCTCTATAGCTCCCAAATCTGCCTCTCGAGAAAC 26908

Qy    239 TATTTATTTCTTCTATACCAGTACCGTACTCTCCCACTGGGATTATTTTAAAATCTCATCT---- 298
          ||||||||||| |||||||||| |||||| ||| ||||||||||| ||||||||||
Db  26907 AAGCTATATTTATTTTTCTATAAAATCTCATCT----TTCTGTCAGGTGTGGTG 26859

Qy    299 CCAGATGACATTTTATCCCTAAATACTTTAGATAAAAGGTGTT 358
          ||||| ||||||| ||||| |||| ||| | |||| |||||||
Db  26858 GCTCACGCCTGTAATCCCAACACTTTGGGAGGTCAGGA 26799

Qy    359 CCTCAGGACCAGCCTGGCCAACACATGGTGAAAACCCTGTCT 418
          |||||| |||||||||| |||||| |||||||||||||||||
Db  26798 GTTTGAGACCAGCCTGACCAACATGGTGAAAACGCAAAAAATTA 26739
```

```
Qy    419  GCTTGGCATGGTCGTGGGCACCTTGTAATCCCAGCTACTTCAGGAAGCTGAGGCAGGAGAAT  478
           ||  ||   |||| |||||||||||||||||||||||| || ||||||||     |||||
Db  26738  GCCGGGTGTGGTGGTGGGCCTCCTGTAATCCCAGCTACTCGGGAGGC---AGAAT  26682

Qy    479  CACTTGAATCCGGGAAGGCAGAGAGATTGCAGTGAGCTGAGATTGCAGTCGAGCCTGGGCGAC  538
           |  |||||   ||| ||  ||||||||||||||||| |              ||    ||
Db  26681  CGATTGAACCTGGGAGGTGGAGGTTGCAGTGAGCCGAGATCACGCCATTGCACTTCAGGC  26622

Qy    539  AGAGACAGAAATGAAACTCTGTCTCAAAAAACAAACAAAAAAAACCACTATACAATAAA  598
           |        |  |   |  | | |  ||||||| |||||||||| | | |
Db  26621  TGGACAGTAAGAGACAAAAACTCCATCTCGGGAAAAACAAAAAACAAAAAAACTCATC  26562

Qy    599  AATGAACAAATGATGCCACAATAGCCACC  625
           |      |   |   | ||  ||| |||
Db  26561  TTTCTAGGTTTTTTAAAAAATTAGAACC  26535
```

OM nucleic - nucleic search, using sw model

```
                    GenCore version 5.1.4_p5_4578
                    Copyright (c) 1993 - 2003 Compugen Ltd.

Run on:             March 29, 2003, 17:03:35 ; Search time 2868.99 Seconds
                    (without alignments)
                    8763.538 Million cell updates/sec

Title:              US-09-691-219-3_COPY_1_1000
Perfect score:      1000
Sequence:           1 cctgccacatgctggcta...........caagaggcctgcacctctga 1000

Scoring table:      IDENTITY_NUC
                    Gapop 10.0 , Gapext 1.0

Searched:           24791104 seqs, 12571243825 residues

Total number of hits satisfying chosen parameters:       49582208

Maximum DB seq length: 2000000000
Minimum DB seq length: 0

Post-processing: Minimum Match 0%
                 Maximum Match 100%
                 Listing first 45 summaries

Database :       Pending_Patents_NA_Main:*
```

```
 1: /cgn2_6/ptodata/1/pna/PCTUS_COMB.seq:*
 2: /cgn2_6/ptodata/1/pna/US06_COMB.seq:*
 3: /cgn2_6/ptodata/1/pna/US07_COMB.seq:*
 4: /cgn2_6/ptodata/1/pna/US080_COMB.seq:*
 5: /cgn2_6/ptodata/1/pna/US081_COMB.seq:*
 6: /cgn2_6/ptodata/1/pna/US082_COMB.seq:*
 7: /cgn2_6/ptodata/1/pna/US083_COMB.seq:*
 8: /cgn2_6/ptodata/1/pna/US084_COMB.seq:*
 9: /cgn2_6/ptodata/1/pna/US085_COMB.seq:*
10: /cgn2_6/ptodata/1/pna/US086_COMB.seq:*
11: /cgn2_6/ptodata/1/pna/US087_COMB.seq:*
12: /cgn2_6/ptodata/1/pna/US088_COMB.seq:*
13: /cgn2_6/ptodata/1/pna/US089_COMB.seq:*
14: /cgn2_6/ptodata/1/pna/US090_COMB.seq:*
15: /cgn2_6/ptodata/1/pna/US091_COMB.seq:*
16: /cgn2_6/ptodata/1/pna/US092_COMB.seq:*
17: /cgn2_6/ptodata/1/pna/US093_COMB.seq:*
18: /cgn2_6/ptodata/1/pna/US094_COMB.seq:*
19: /cgn2_6/ptodata/1/pna/US095A_COMB.seq:*
20: /cgn2_6/ptodata/1/pna/US095B_COMB.seq:*
21: /cgn2_6/ptodata/1/pna/US095C_COMB.seq:*
22: /cgn2_6/ptodata/1/pna/US095D_COMB.seq:*
23: /cgn2_6/ptodata/1/pna/US096A_COMB.seq:*
24: /cgn2_6/ptodata/1/pna/US096B_COMB.seq:*
25: /cgn2_6/ptodata/1/pna/US096C_COMB.seq:*
26: /cgn2_6/ptodata/1/pna/US096D_COMB.seq:*
27: /cgn2_6/ptodata/1/pna/US096E_COMB.seq:*
28: /cgn2_6/ptodata/1/pna/US097A_COMB.seq:*
29: /cgn2_6/ptodata/1/pna/US097B_COMB.seq:*
30: /cgn2_6/ptodata/1/pna/US097C_COMB.seq:*
31: /cgn2_6/ptodata/1/pna/US098A_COMB.seq:*
32: /cgn2_6/ptodata/1/pna/US098B_COMB.seq:*
33: /cgn2_6/ptodata/1/pna/US098C_COMB.seq:*
34: /cgn2_6/ptodata/1/pna/US099A_COMB.seq:*
35: /cgn2_6/ptodata/1/pna/US099B_COMB.seq:*
36: /cgn2_6/ptodata/1/pna/US099C_COMB.seq:*
37: /cgn2_6/ptodata/1/pna/US099D_COMB.seq:*
38: /cgn2_6/ptodata/1/pna/US100A_COMB.seq:*
39: /cgn2_6/ptodata/1/pna/US100B_COMB.seq:*
40: /cgn2_6/ptodata/1/pna/US101A_COMB.seq:*
41: /cgn2_6/ptodata/1/pna/US101B_COMB.seq:*
42: /cgn2_6/ptodata/1/pna/US102A_COMB.seq:*
43: /cgn2_6/ptodata/1/pna/US102B_COMB.seq:*
44: /cgn2_6/ptodata/1/pna/US6000_COMB.seq:*
45: /cgn2_6/ptodata/1/pna/US6001_COMB.seq:*
46: /cgn2_6/ptodata/1/pna/US6002_COMB.seq:*
47: /cgn2_6/ptodata/1/pna/US6003_COMB.seq:*
48: /cgn2_6/ptodata/1/pna/US6004_COMB.seq:*
49: /cgn2_6/ptodata/1/pna/US6005_COMB.seq:*
50: /cgn2_6/ptodata/1/pna/US6006_COMB.seq:*
51: /cgn2_6/ptodata/1/pna/US6007_COMB.seq:*
52: /cgn2_6/ptodata/1/pna/US6008_COMB.seq:*
53: /cgn2_6/ptodata/1/pna/US6009_COMB.seq:*
54: /cgn2_6/ptodata/1/pna/US6010_COMB.seq:*
55: /cgn2_6/ptodata/1/pna/US6011_COMB.seq:*
56: /cgn2_6/ptodata/1/pna/US6012_COMB.seq:*
57: /cgn2_6/ptodata/1/pna/US6013_COMB.seq:*
58: /cgn2_6/ptodata/1/pna/US6014_COMB.seq:*
59: /cgn2_6/ptodata/1/pna/US6015_COMB.seq:*
60: /cgn2_6/ptodata/1/pna/US6016_COMB.seq:*
61: /cgn2_6/ptodata/1/pna/US6017_COMB.seq:*
62: /cgn2_6/ptodata/1/pna/US6018_COMB.seq:*
63: /cgn2_6/ptodata/1/pna/US6019_COMB.seq:*
64: /cgn2_6/ptodata/1/pna/US6020_COMB.seq:*
65: /cgn2_6/ptodata/1/pna/US6021_COMB.seq:*
66: /cgn2_6/ptodata/1/pna/US6022_COMB.seq:*
67: /cgn2_6/ptodata/1/pna/US6023_COMB.seq:*
68: /cgn2_6/ptodata/1/pna/US6024_COMB.seq:*
69: /cgn2_6/ptodata/1/pna/US6025_COMB.seq:*
70: /cgn2_6/ptodata/1/pna/US6026_COMB.seq:*
71: /cgn2_6/ptodata/1/pna/US6027_COMB.seq:*
72: /cgn2_6/ptodata/1/pna/US6028_COMB.seq:*
73: /cgn2_6/ptodata/1/pna/US6029_COMB.seq:*
74: /cgn2_6/ptodata/1/pna/US6030_COMB.seq:*
75: /cgn2_6/ptodata/1/pna/US6031_COMB.seq:*
76: /cgn2_6/ptodata/1/pna/US6032_COMB.seq:*
77: /cgn2_6/ptodata/1/pna/US6033_COMB.seq:*
78: /cgn2_6/ptodata/1/pna/US6034_COMB.seq:*
79: /cgn2_6/ptodata/1/pna/US6035_COMB.seq:*
80: /cgn2_6/ptodata/1/pna/US6036_COMB.seq:*
81: /cgn2_6/ptodata/1/pna/US6037_COMB.seq:*
82: /cgn2_6/ptodata/1/pna/US6038_COMB.seq:*
83: /cgn2_6/ptodata/1/pna/US6039_COMB.seq:*
84: /cgn2_6/ptodata/1/pna/US6040_COMB.seq:*
85: /cgn2_6/ptodata/1/pna/US6041_COMB.seq:*
```

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	394.6	39.5	438	19	US-09-528-409-19572	Sequence 19572, A
2	394.6	39.5	438	35	US-09-933-524-19572	Sequence 19572, A
3	394.6	39.5	438	35	US-09-933-524A-19572	Sequence 19572, A
4	302	30.2	449	19	US-09-528-409-18017	Sequence 18017, A
5	302	30.2	449	35	US-09-933-524-18017	Sequence 18017, A
6	302	30.2	449	35	US-09-933-524A-18017	Sequence 18017, A
7	269.8	27.0	557	62	US-60-182-316-7828	Sequence 7828, Ap
8	231.8	23.2	557	68	US-60-243-468-131	Sequence 131, App
9	227.2	22.7	35934	64	US-60-207-317-39	Sequence 39, Appl
10	226	22.6	10702	30	PCT-US01-01354-24569	Sequence 24569, A
11	226	22.6	10702	30	PCT-US01-01354-24569	Sequence 24569, A
12	226	22.6	10702	39	US-09-764-905-24569	Sequence 24569, A
13	222.4	22.2	1503	36	US-10-092-399-24569	Sequence 780, App
14	217.6	21.8	1932	36	US-09-973-278-780	Sequence 9471, Ap
15	217.6	21.8	1932	40	US-09-950-083-9471	Sequence 11622, A
16	217.6	21.8	59065	40	US-10-105-299-11622	Sequence 11622, A
17	217.6	21.8	10702	30	US-10-135-696-3	Sequence 3, Appli
18	217.6	21.8	128978	1	PCT-US01-17423-6	Sequence 6, Appli
19	217.6	21.8	128978	1	PCT-US02-17423-6	Sequence 6, Appli
20	217.6	21.8	128978	41	PCT-US02-25766-4301	Sequence 4301, Ap
21	217.4	21.7	32768	65	US-10-161-453-6	Sequence 6, Appli
					US-60-212-358-62	Sequence 62, Appl

```
c  22  216.2  21.6  146776  68  US-60-248-505-474    Sequence 474, Appl
c  23  216    21.6    8338  68  US-60-245-222-39     Sequence 39, Appli
c  24  215.2  21.5  163350  30  US-60-793-706A-3     Sequence 3, Appli
c  25  214.6  21.4   32768  65  US-09-213-178-285    Sequence 285, Appl
c  26  213.6  21.4   32768  64  US-60-207-416-36     Sequence 36, Appl
c  27  213.2  21.3  131943  66  US-60-226-176-1562   Sequence 1562, Ap
c  28  213.2  21.3  131943  67  US-60-233-468-1562   Sequence 1562, Ap
c  29  213.2  21.3  131943  75  US-60-313-371-1562   Sequence 1562, Ap
c  30  213    21.3   32768  66  US-60-229-525-101    Sequence 101, App
c  31  213    21.3   32768  66  US-60-229-525-203    Sequence 203, Appli
c  32  212.6  21.3   74637  68  US-60-243-468-9      Sequence 9, Appli
c  33  212.6  21.3  103367  68  US-60-242-679-372    Sequence 372, Appl
c  34  211.6  21.3   19212  65  US-60-213-800-156    Sequence 156, Appl
c  35  211    21.1   13216   1  PCT-US01-01329-2666  Sequence 2666, Ap
c  36  211    21.1   13216  30  PCT-US01-01339-7810  Sequence 7810, Ap
c  37  211    21.1   13216  30  US-09-764-891-7810   Sequence 7810, Ap
c  38  211    21.1   13216  39  US-10-080-090-2666   Sequence 2666, Ap
c  39  211    21.1   13216  42  US-10-205-303-2666   Sequence 2666, Ap
c  40  210.2  21.0  170245  66  US-60-226-176-1778   Sequence 1778, Ap
c  41  210.2  21.0  170245  67  US-60-233-468-1778   Sequence 1778, Ap
c  42  210.2  21.0  170245  75  US-60-113-371-1778   Sequence 1778, Ap
c  43  209.4  20.9   32768  68  US-60-243-742-79     Sequence 79, Appli
c  44  209.4  20.9   18873  70  US-60-262-080-1      Sequence 1, Appli
c  45  209.4  20.9   29163   1  PCT-US01-01329-2665  Sequence 2665, Ap
```

ALIGNMENTS

RESULT 1
US-09-528-409-19572/c
; Sequence 19572, Application US/0952840

US-09-528-409-19572

Query Match   39.5%; Score 394.6; DB 19; Length 438;
Best Local Similarity   99.0%; Pred. No. 2.8e-42;
Matches   397; Conservative   0; Mismatches   4; Indels   0; Gaps   0;

```
Qy  600  ATGAACAATGATGCCACAATAGCACCAGAGAATTTTATAAATACACAGAATTCCCAGGCCCTG  659
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  438  ATGAACAATGATGCCACAATAGCACCAGAGAATTTTATAAATACACAGAATTCCCAGGCCCTG  379

Qy  660  CCCCAGACCTACTGAATCCTGGAAATATTCAGGCTCCACACCCAGAGATTCTGGTTCGGT  719
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  378  CCCCAGACCTACTGAATCCTGGAAATATTCAGGCTCCACACCCAGAGATTCTGGTTCGGT  319

Qy  720  TGGTCTGATGCAGGGACCTGTTAACATCCTGGAAATATTCAGGCTTAATGCTGAGCC  779
         |||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  318  TGGTCTGATGCAGGGACCTGTTAACATCCTGGAAATATTCAGGCTTAATGCTGAGCC  259

Qy  780  TGGTCTGGTGCTCAGAGAGTAGACAGAGACCTGGAGAAAAACCAGGGTGTCTGAGGTTTTCCAGAAG  839
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  258  TGCTCGGTGCTCAGAGAGTAGACAGAGACCTGGAGAAAAACCAGGGTGTCTGAGGTTTTCCAGAAG  199
```

RESULT 2
US-09-933-524-19572/c
; Sequence 19572, Application US/0993352

US-09-933-524-19572

Query Match   39.5%; Score 394.6; DB 35; Length 438;
Best Local Similarity   99.0%; Pred. No. 2.8e-42;
Matches   397; Conservative   0; Mismatches   4; Indels   0; Gaps   0;

```
Qy  600  ATGAACAATGATGCCACAATAGCACCAGAGAATTTTATAAATACACAGAATTCCCAGGCCCTG  659
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  438  ATGAACAATGATGCCACAATAGCACCAGAGAATTTTATAAATACACAGAATTCCCAGGCCCTG  379

Qy  660  CCCCAGACCTACTGAATCCTGGAAATATTCAGGCTCCACACCCAGAGATTCTGGTTCGGT  719
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  378  CCCCAGACCTACTGAATCCTGGAAATATTCAGGCTCCACACCCAGAGATTCTGGTTCGGT  319

Qy  720  TGGTCTGATGCAGGGACCTGTTAACATCCTGGAAATATTCAGGCTTAATGCTGAGCC  779
         |||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  318  TGGTCTGATGCAGGGACCTGTTAACATCCTGGAAATATTCAGGCTTAATGCTGAGCC  259

Qy  780  TGCTGGGTGCTCAGAGTAGACAGGGACCTGGAGAAAAACCAGGGTGTCTGAGGTTTTCCAGAAG  839
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  258  TGCTGGGTGCTCAGAGTAGACAGGGACCTGGAGAAAAACCAGGGTGTCTGAGGTTTTCCAGAAG  199

Qy  840  AAAACCAGAGTCCAGAGAAGCAGAGAGCCACTCAGTGAGGACCCAAGCAGAGACTGAAAGAGAGCGGGTGCA  899
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  198  AAAACCAGAGTCCAGAGAAGCAGAGAGCCACTCAGTGAGGACCCAAGCAGAGACTGAAAGAGAGCGGGTGCA  139

Qy  900  CCTCACATCCTCACTCCTGGCACCCCGTGTCTCCTACAAGATGAGAGGAGGACTGAAAGAGAGCCCCT  959
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  138  CCTCACATCCTCACGCCCTGGCACCCCGTGTCTCCTACAAGATGAGAGGAGGACTGAAAGAGAGCCCCT  79

Qy  960  TCCTGTCCCCAGTGGTGTGGGCAAGAGAGGCCTGCACCTCTGA  1000
         |||||||||||||||||||||||||||||||||||||||||||
Db  78   TCCTGTCCACAGTGGTCTGGGCAAGAGAGGCCTGCACCTCTGA  38
```

RESULT 3
US-09-933-524A-19572/c
; Sequence 19572, Application US/0993352A

GENERAL INFORMATION:
; APPLICANT: Drmanac, Radoje T.
; APPLICANT: Labat, Ivan
; APPLICANT: Stache-Crain, Birgit
; APPLICANT: Dickson, Mark
; APPLICANT: Jones, Lee W.
; TITLE OF INVENTION: Novel Nucleic Acid Sequences Obtained
; TITLE OF INVENTION: From Various Libraries
; FILE REFERENCE: 774
; CURRENT APPLICATION NUMBER: US/09/933,524A
; CURRENT FILING DATE: 2001-08-20
; PRIOR APPLICATION NUMBER: 09/528,409
; PRIOR FILING DATE: 2000-05-17
; NUMBER OF SEQ ID NOS: 116231
; SOFTWARE: Hy-patent.pl Version 3.1
; SEQ ID NO 19572
; LENGTH: 438
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-933-524A-19572

Query Match   39.5%; Score 394.6; DB 35; Length 438;
Best Local Similarity 99.0%; Pred. No. 2.8e-42;
Matches 397; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Qy   600 ATGAACAATGATGCCACAATAGGACACCAGAGAATTTATAAAATACAGAATTCCCAGGCCTG 659
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   438 ATGAACAATGATGCCACAATAGGACACCAGAGAATTTTATAAAATACAGAATTCCCAGGCCTG 379

Qy   660 CCCCAGACCTACTGAATCCTGGAAATATTCAGCGAGATTCTCGTTCCGT 719
          |||||||||||||||||||||||||||||||||||||||||||||||||
Db   378 CCCCAGACCTACTGAATCCTGGAAATATTCAGCGAGATTCTCGTTCGGT 319

Qy   720 TGGTCTGAATGCAGGGACCTGTAACCTCGGTTGTAACACCTTCCACACCCAGAGAATTCAGC 779
          |||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   318 TGGTCTGATTGCAGGGACCTGTAACCTCGGTTGTAACACCTTCCAGTGTAATGCTGAGCC 259

Qy   780 TGCTGGTGCTCAGAGTAGACACAGACCTGGAGAAAACCAGGGTGTCTGAGGTTTTCCAGAAG 839
          |||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   258 TGCTGGTGCTCAGAGTAGACACAGACCTGGAGAAAACCAGGGTGTCTGAGGTTTTCCAGAAG 199

Qy   840 AAAACCAGAGTCCAGAGAAGCAGAGAGGCCACTCAGTGAGGACCCAAGCACGAGAGCGGGTGCA 899
          |||||||||||||||||||||||||||||||||||||||||||||||||
Db   198 AAAACCAGAGTCCAGAGAAGCAGAGAGGCCTTCAGTGAGGACCCAAGCACGAGAGCGGGTGCA 139

Qy   900 CCTCACATCCTCACTCCTGGCACCCCCTTCTCCTACAGATGAGGACTGAAAGGAGAGCCCCT 959
          ||||||||||||||||||||||||||||||||||||||||||||||||
Db   138 CCTCACATCCTCACGCCTGGCACCCCGTCCTCCTACACAGATGAGGACTGAAAGAGAGCCCCT 79

Qy   960 TCCTGTGTCCCAGTGGTGTGGGCAAGAGAGGCCTGCACCTCTGA 1000
          |||||||||||||||||||||||||||||||||||||||||||
Db   78 TCCTGTCCACAGTGGTGTCGGGCAAGAGAGGCCTGCACCTCTGA 38

RESULT 4
US-09-528-409-18017/c
Sequence 18017, Application US/09528409
; GENERAL INFORMATION:
; APPLICANT: Drmanac, Radoje T.
; APPLICANT: Labat, Ivan
; APPLICANT: Stache-Crain, Birgit
; APPLICANT: Dickson, Mark
; APPLICANT: Jones, Lee W.
; TITLE OF INVENTION: Novel Nucleic Acid Sequences Obtained
; TITLE OF INVENTION: From Various Libraries
; FILE REFERENCE: 774
; CURRENT APPLICATION NUMBER: US/09/528,409
; CURRENT FILING DATE: 2000-03-17
; PRIOR APPLICATION NUMBER: 60/125,453
; PRIOR FILING DATE: 1999-03-19
; NUMBER OF SEQ ID NOS: 116231
; SOFTWARE: Hy-patent.pl Version 3.1
; SEQ ID NO 18017

; LENGTH: 449
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-528-409-18017

Query Match   30.2%; Score 302; DB 19; Length 449;
Best Local Similarity 84.3%; Pred. No. 2.9e-30;
Matches 364; Conservative 0; Mismatches 65; Indels 3; Gaps 2;

Qy   340 TCTTTGAAAAAAAANCATAAACCTCAGGACCCAGCCTGGCCAACATGG--TGAAACCCTGTGTCT 397
          || ||| |||||||| ||||||||||||||||||||||||||||||||| | | | | |
Db   447 TTTTGAATAGGAATCATACCGTCAGGACGAGCCCCCCCAAACATGTTTTAACAACGGCC 388

Qy   398 GTACTAAAAATAACAAAAAATTAGCTTGGCATGGTGGTCGTGGGGCACCTGTAAATCCCAGCTACTC 457
          | || ||| | |||||||||||||||||||||||||||| ||||| ||||| ||||||
Db   387 GCCCTAAAAATAACAAAAAATTAGCTTGGCATGGTGGTGTAAATCTCAGTTATAGC 328

Qy   458 AGGAAGCTGAGGCAGGAG-AATCACTTGAATCCGGGAAGCAGAGAGTTGCAGTGAGCTGAG 516
          ||||| |||||||||||| | | | | ||||||||||||||||||||||||||
Db   327 TAGGCCTTGAGGTCTGTAGGCGTTCGCTTGAATCCGGGAAGCAGAGAGATTGCAGTGAGCTGAG 268

Qy   517 ATTGCAGTCCAGCCCTGGGGCGACAGAGACAGAAAATAAAACTCTGTCTCAAAAACAAACAAA 576
          ||| ||||||| | ||||| |||||||||||||||||||||||||||||||||||||||
Db   267 ATTGCAGTCCAGCCCTGGAGTTACAGAGACAGAAAATGAACCTGTGTCTCAAACAACGACCAAA 208

Qy   577 CAAAAAAAAACCACTATACATAAAAATGAACAAATGAATGATGCCACAGAAGAATTTTA 636
          |||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   207 CAAAAAAAACCCCTATTCATTAAAAATGAACAAATGATGATGCCACAATAGGACACCAGAGAATTTTA 148

Qy   637 TAAAATACAGAATTCCCAGGCCCTGCCCCAGACCTACTGAATCCTGGAAATATTCAGCGAGATTCAGGCTCC 696
          ||||||||||||||||| |||||||||||||||||||||||||||||||||||
Db   147 TAAAATACAGAATTCCCAGGCCTGCCCCAGACCTACTGAATCCTGGAAATATTCAGCGAGATTCTCC 88

Qy   697 ACACCCAGAGAATTCTGGTTCGGTTGGTCTGAATGCAGGGACCTGTAACCTCGGTTGTAACA 756
          ||||| |||||| ||||||||||||||||||||||||||||||||||||
Db   87 ACACCCAGAGATTTCTGGTTCGGTTGGTCTGATTGCAGGGACCTGTAACCTCGGTTGTAACA 28

Qy   757 CCTTCCCAGGT 768
          ||| || ||
Db   27 CCTCCCACGGT 16

RESULT 5
US-09-933-524-18017/c
Sequence 18017, Application US/09933524
; GENERAL INFORMATION:
; APPLICANT: Drmanac, Radoje T.
; APPLICANT: Labat, Ivan
; APPLICANT: Stache-Crain, Birgit
; APPLICANT: Dickson, Mark
; APPLICANT: Jones, Lee W.
; TITLE OF INVENTION: Novel Nucleic Acid Sequences Obtained
; TITLE OF INVENTION: From Various Libraries
; FILE REFERENCE: 774
; CURRENT APPLICATION NUMBER: US/09/933,524
; CURRENT FILING DATE: 2001-08-20
; PRIOR APPLICATION NUMBER: 09/528,409
; PRIOR FILING DATE: 2000-05-17
; NUMBER OF SEQ ID NOS: 116231
; SOFTWARE: Hy-patent.pl Version 3.1
; SEQ ID NO 18017
; LENGTH: 449
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-933-524-18017

Query Match   30.2%; Score 302; DB 35; Length 449;
Best Local Similarity 84.3%; Pred. No. 2.9e-30;
Matches 364; Conservative 0; Mismatches 65; Indels 3; Gaps 2;

Qy   340 TCTTTGAAAAAAAANCATAAACCTCAGGACCCAGCCTGGCCAACATGG--TGAAACCCTGTGTCT 397
          || ||| |||||||| ||||||||||||||||||||||||||||||||| | | | | |
Db   447 TTTTGAATAGGAATCATACCGTCAGGACGAGCCCCCCCAAACATGTTTTAACAACGGCC 388

```
Qy   398  GTACTAAAAATACAAAAATTAGCTTGGCATGGTCGTGGGCACCTGTAATCCAGCTACTC  457
          ||||   |||||||||||||| | |  |||| | ||| |      ||||||||||||| |||
Db   387  GCCCTAAAAATACAAAAATTCATAACCTGGCATGGTAGGCGAGCTTCTTGTAATTCAGTGTATAGC  328

Qy   458  AGGAAGCTGAGGCAGGAG-AATCACTTGAATCCGGGAAGCAGAGAATTGCAGTGAGCTGAG  516
          | |||  | |||   |||  | |||||||||||||||||||||||||||||||||||||||
Db   327  TAGGCCTTGAGTCTGTAGCGTTCGCTTGAATCCGGGAAGCAGAGAATTGCAGTGAGCTGAG  268

Qy   517  ATTGCAGTCGAGCCTGGGACGACAGAGACAGAAATGAAACTCTGTCTCAAAAACAAACAA  576
          |||||||||||||||||||  | ||||||||||||||||| ||||||||||| ||| ||||
Db   267  ATTGCAGTCGAGCCTGGGAGTACAGAGACAGAAATGAAACCTGTCTCAAAAGACGAGCAAA  208

Qy   577  CAAAAAAACCACTATACATAAAAATGAACAATGACACCAGAGAATTTTA  636
          |||||||| |  |||  |||||||||||||||||   |||||||||| |
Db   207  CAAAAAAACCCTATTCATAAAAATGAACAATGATGCCACAATAGCACCAGAGAATTTA  148

Qy   637  TAAATACAGATTCCCAGGCGCCTGCCCCAGACCTACTGAATCCTGGAAATATTCAGGCTCC  696
          ||||||||||||||||||||||||||||||| ||||||||||||||||||||||||||||||
Db   147  TAAATACAGATTCCCAGGCGCCTGCCCCAGACTTACTGAATCCTGGAAATATTCAGGCTCC  88

Qy   697  ACACCCAGAGATTCTGGTTCGGTTCTGATGCAGGGAGCCCTGTAACCTGCGTTGTTCAGGCTGTTCAGGCTGTTAACA  756
          |||||||||||||| |||||||||  |||||||||  |||||||||||||||||||||  |||
Db    87  ACACCCAGAGATTCTCGGTTCGGTTGGTCTCGAATGCAGGAACCTGTAACCTGCGTTGTTAACA  28

Qy   757  CCTTCTCCAGGT  768
          ||  | |||
Db    27  CCTCCCACGCGT  16
```

RESULT 6
US-09-933-524A-18017/c
; Sequence 18017, Application US/0933524A
; GENERAL INFORMATION:
; APPLICANT: Drmanac, Radoje T.
; APPLICANT: Labat, Ivan
; APPLICANT: Stache-Crain, Birgit
; APPLICANT: Dickson, Mark
; APPLICANT: Jones, Lee W.
; TITLE OF INVENTION: Novel Nucleic Acid Sequences Obtained
; TITLE OF INVENTION: From Various Libraries
; FILE REFERENCE: 774
; CURRENT APPLICATION NUMBER: US/09/933,524A
; CURRENT FILING DATE: 2001-08-20
; PRIOR APPLICATION NUMBER: 09/528,409
; PRIOR FILING DATE: 2000-05-17
; NUMBER OF SEQ ID NOS: 116231
; SOFTWARE: Hy-patent.pl Version 3.1
; SEQ ID NO 18017
; LENGTH: 449
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-933-524A-18017

Query Match   Best Local Similarity   84.3%;
Matches   364;   Conservative   0;   Mismatches   65;   Indels   3;   Gaps   2;
Score 302;   DB 35;   Length 449;

```
Qy   340  TCTTTGAAAAAATCATAACCTCAGGACCAGCCTGGCCAACATGG--TGAAACCCTGTCT  397
          ||  |  |   | | |    |   ||  | ||||  ||||| ||  |||||||||||||
Db   447  TTTTGAATAGGAATCATACCGTCAGGACGAGCCCCCAAACATGTTTTAACAACGGCCC  388

Qy   398  GTACTAAAAATACAAAAATTAGCTTGGCATGGTGTCGTGGGCACCTGTAATCCAGCTACTC  457
          ||  |  |||||||||| | |  | ||  |   ||| |   |||||||||||||||||||||
Db   387  GCCCTAAAAATACAAAAATTACTTGGCATGGTAGCCAGCTCTTGTAATCCCAGCACTTTGG  328

Qy   458  AGGAAGCTGAGGCAGGAG-AATCACTTGAATCCGGGAAGGCAGAGATTGCAGTGAGCTGAG  516
          | |||  | |||   |||  | |||||||||||||||||||||||||||||||||||||||
Db   327  GCCCTAAAAATACAAAAATTACTTGGCATGGTGTCGCATGCCTGTAATCCCAGCTACTC  268

Qy   517  ATTGCAGTCGAGCCCTGGGCGACAGAGACGAAAATGAACTCTGTCTCAAAAACGAGCAAA  576
          |||||||||||||  || | ||||||||| ||||||||  ||||||||||| | | |||
Db   267  TAGGCCTTGAGTCGAGCGGTTCGCTTGAATCCGGGAAGCAGAGATTGCAGTGAGCTGAG  208
```

RESULT 7
US-60-182-316-7828
; Sequence 7828, Application US/60182316
; GENERAL INFORMATION:
; APPLICANT: Curtis, Anne L.
; APPLICANT: Lagace, Robert E.
; APPLICANT: Klingler, Tod M.
; APPLICANT: Stuve, Laura L.
; TITLE OF INVENTION: CpG Island Polynucleotides
; FILE REFERENCE: PX-0003 P
; CURRENT APPLICATION NUMBER: US/60/182,316
; CURRENT FILING DATE: 2000-02-10
; NUMBER OF SEQ ID NOS: 14,630
; SOFTWARE: PERL Program
; SEQ ID NO 7828
; LENGTH: 557
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; OTHER INFORMATION: Incyte ID NO: TVD30260337.R2
; FEATURE:
; NAME/KEY: unsure
; LOCATION: 89-150, 152-259
; OTHER INFORMATION: a, t, c, g, or other
US-60-182-316-7828

Query Match   Best Local Similarity   61.2%;
Matches   271;   Conservative   0;   Mismatches   172;   Indels   0;   Gaps   0;
Score 269.8;   DB 62;   Length 557;

```
Qy   577  CAAAAAAACCACTATACATAATAAAAATGAACAATGATGCCACAATAGCACCAGAGAATTTTA  636
          ||||||||||  |||  |||  |||||||||||||||||||||||||||||||||||||||| |
Db   207  CAAAAAAACCCTATTCATAAAAATGAACAATGATGCCACAATAGCACCAGAGAATTTA  148

Qy   637  TAAATACAGATTCCCAGGCGCCTGCCCCAGAGATTCTGGTTTCGGTTGGTCAATATTCAGGCTCC  696
          ||||||||||||||||||||||||||||||  | | | |  || ||| |||||||||||||||||
Db   147  TAAATACAGATTCCCAGGCGCCTGCCCCAGACTTACTGAATCCTGGAAATATTCAGGCTCC  88

Qy   697  ACACCCAGAGATTCTGGTTCGGTTGGTCGGTTGATGCAGGGACCTGTAACCTGCGTTGTTAACA  756
          |||||||||||||| |||||||||||||  || |||||||  |||||||||||||||||||||||
Db    87  ACACCCAGAGATTCTCGGTTCGGTTGGTCTCGAATGCAGGAACCTGTAACCTGCGTTGTTAACA  28

Qy   757  CCTTCTCCAGGT  768
          ||  | |||
Db    27  CCTCCCACGCGT  16

Qy   558  TGTCTCAAAAACAAACAAACAAAAAAAACCACTATACATAAAAATGAACAATGATGCCACA  617
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    31  TGTCTCAAAAACAAACAAACAAAAAAAACCACTATACATAAAAATGAACAATGCCANN  90

Qy   618  ATAGCACCAGAGAATTTTATATAAAATACAGATTCCCAGGCCCTGCCCCAGACCTACTGAATC  677
          ||| ||||||||||||||||||
Db    91  NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN  150

Qy   678  CTGGAAATATTCAGGCTCCACACCCAGAGATTCTGGTTTCGGTTGGTCTCGATGCAGGGACC  737
          |
Db   151  CNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN  210

Qy   738  TGTAACCTGCGTTGTTAACACCTTCTCCAGGTAATGCTGAGCCTGGTGGCTCAGAGTAG  797
          |
Db   211  NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNGCTCAGAGTAG  270

Qy   798  ACAGACCTGGAGAAAACCAGGGGTGTCTGAGGTTTTTCCAGAAGAAAAACAGAGTCCAGA  857
          ||||||||||||||||||||||||||||||||||||||||||||||||||| |||||||||
Db   271  ACAGACCTGGAGAAAACCAGGGTGTCTGAGGTTTTTCCAGAAGAAAAACAGAGTCCAGA  330

Qy   858  AGCAGAGAGGCACTCAGTGAGGACCCAAGCAGACGAGGGGTGCACCTCACACTCCTCACT  917
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   331  AGCAGAGAGGCACTCAGTGAGGACCCAAGCAGACGAGGGGTGCACCTCACATCCTCACT  390
```

```
Qy   918 GGCACCCCGTCTCCTACAAGATGAGAGAGACTGAAAGAGCCCCTTCCCTGTGGTGT 977
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   391 GGCACCCCGTCTCCTACAAGATGAGAGAGACTGAAAGAGCCCCTTCCCTGTGGTCT 450

Qy   978 GGGCAAGAGGGCCTGCACCTCTGA 1000
         ||||||||||||||||||||||||
Db   451 GGGCAAGAGGGCCTGCACCTCTGA 473
```

RESULT 8
US-60-243-468-131
Sequence 131, Application US/60243468
; GENERAL INFORMATION:
; APPLICANT: Beasley, Ellen
; TITLE OF INVENTION: ISOLATED HUMAN DRUG TARGET PROTEINS,
; TITLE OF INVENTION: NUCLEIC ACID MOLECULES ENCODING HUMAN DRUG TARGET PROTEINS,
; TITLE OF INVENTION: AND USES THEREOF
; FILE REFERENCE: CL000929
; CURRENT APPLICATION NUMBER: US/60/243,468
; CURRENT FILING DATE: 2000-10-27
; NUMBER OF SEQ ID NOS: 2121
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 131
; LENGTH: 35934
; TYPE: DNA
; ORGANISM: HUMAN
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)...(35934)
; OTHER INFORMATION: n = A,T,C or G
US-60-243-468-131

Query Match   23.2%;   Score 231.8;   DB 68;   Length 35934;
Best Local Similarity   66.3%;   Pred. No. 1.5e-21;
Matches   430;   Conservative   0;   Mismatches 202;   Indels   17;   Gaps   6;

```
Qy     1 CCTGCCACCATGCCCTGCCTAATTTTCTTATTTTAGTAGAGACGGAGGTTTTGCCATGTTG 60
           |||||||||| ||| |||  ||||||||| ||| ||| |||||||| ||| |||||||| |
Db 14023 CCTGCCACCACTGGCTAGTTTTTGTAGTTCTAGTAGAGACAGGGTTTCACCATGGTG 14082

Qy    61 ACCAGGGCTGGTCTCGAACTCCTCAGGTGATCC-GCCTGCCTCAGCCTCCCAAAGT 119
           || ||| |||| |||||| |||||||| || || |||||| ||||||||||||||||
Db 14083 GCCAGGCTGGTCTCGAACTCCTGGACCTCAGGTGATCCAGCCTCCCAAAGT 14142

Qy   120 GCTGGGATTATAGGCGTGCAGCCCCGCCACCCAACACATTTTTTAAATACTGAAAGTA 179
           ||||||| |||||| ||||||||| ||||| |||||||||| || |||||||||| |
Db 14143 GCTGGGTTTACAGGCATGAGCCCACCACCATGCCCGGC---CCTGTTTTCCATCTTTCAAGGTA 14199

Qy   180 GAGGGAATAGTTATAGTGTACCCCATTTACCCATCACTCAGTTCACAGCGTGGTGACAT 239
           |||||| ||||||| |||||||| ||||| ||| |||||||| |||||||| ||||||
Db 14200 CTGAGAATATATA-----ATGCCATCGTTCCCCACCCTTACACTTCTTCAGTCTCTGGGCATA 14253

Qy   240 ATTTATTTCTTCTATACCAGTACCGTACTCTCCCCACTGGGGATTATTTTTAAGGCAAAACC 299
           || ||  ||  ||||| ||||| ||||||| ||||||||||||| ||||| |||||||||
Db 14254 AAAACTTC-----TCTAAGCACCCTCCATAGATCCAGCCGTGGTG 14308

Qy   300 CAGATGACATTTTATCCCTAAATACTTTGATAAAAGGTGTTCTTTGGAAAAAAATCATAAC 359
           || | |||||||||| |||||||||| |||||||||||||| |||||||||||| |||
Db 14309 GCTCACGCCTGTRAATCCCAGCACTTGGGAGGCCGAGGCAGGTCAGGAG 14368

Qy   360 CTCAGGACCACCAGCCTGGGCCAACAAAACCTGTCTGTCTTGTACTATACAAAAAATTTAG 419
           || | || ||||||| |||||||| ||| |||||| || |||||| | ||||||||||| |
Db 14369 TTTGAGACCAGCCTGGCCAACATGGTGAAAACCCCATCTCTACTAAAAATACAAAAATTAG 14428

Qy   420 CTTGGCCAAGGTGCGTGGGCACCTGTAATCCCAGCTACTCAGGGAGGCAGGAGAATC 479
           ||||| || |||||||||| ||| |||||||||||| |||||||||||||||||||
Db 14429 CTGGGCGTGGTGGCGGGCGCCTGTAATCCCAGCTACTCGGGGAGGCTGAGGCAGGAGAATC 14488

Qy   480 ACTTGAATCCGGGAAGCAGAGATTGCAGTGAGTCGAGCCTGGGCGGACA 539
           ||||||||| || ||||||| | |||| ||| ||| ||||||||| |||
Db 14489 GCTTGAACCTGGGAGGCGGAGGTTGCAGTGAGTTGAGATTGTGAACCCTGCACTCCAGTCT 14548

Qy   540 GAGACAGAAATGAA-ACTCTGTCTCAAAAAACAAACAAAAAAAAAC-CACTATACATAA 597
           |||| || | |||   | || || |||||||| ||| ||||||||| |||| |||| ||
Db 14549 GGGTGACAGAGCAAGAACTCCGTCTCAAAAAAAAAAAAAAAAGAATCACATTTTAGCTAG 14608

Qy   598 GGGTGACAGAGCAAGAACTCCGTCTCAAAAAAAAAAAAAAAAGAATCACATTTTAGCTAG 646
           ||||| ||| ||| |||||||| |||||||||||||||||| |||||| || | | |||
Db 14609 AGAGGAAAAAATGGTGGCAGAAGAACACAGATAGCATATATTAAAGTCAGA 14657
```

RESULT 9
US-60-207-317-39/c
Sequence 39, Application US/60207317
; GENERAL INFORMATION:
; APPLICANT: Beasley, Ellen
; TITLE OF INVENTION: ISOLATED HUMAN G-PROTEIN COUPLED
; TITLE OF INVENTION: RECEPTORS, NUCLEIC ACID MOLECULES ENCODING HUMAN GPCR
; TITLE OF INVENTION: PROTEINS, AND USES THEREOF
; FILE REFERENCE: CL000609
; CURRENT APPLICATION NUMBER: US/60/207,317
; CURRENT FILING DATE: 2000-05-30
; NUMBER OF SEQ ID NOS: 406
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 39
; LENGTH: 18766
; TYPE: DNA
; ORGANISM: HUMAN
US-60-207-317-39

Query Match   22.7%;   Score 227.2;   DB 64;   Length 18766;
Best Local Similarity   65.8%;   Pred. No. 6.6e-21;
Matches   395;   Conservative   0;   Mismatches 193;   Indels   12;   Gaps   4;

```
Qy     1 CCTGCCACCATGCCCTGCCTAATTTTCTTATTTTAGTAGAGACGGAGGTTTTGCCATGTTG 60
           ||||||||||||| |||||||||||||||||||||||||||||||||||||||||||||
Db  9880 CGTGCCACCACACCTGGCTAATTTTGTATTTTTAGTAGAGATGGGGTTTGCCATGTTG 9821

Qy    61 ACCAGGGCTGTCTCGAACTCTCTTGACCTCAGGTGATCCGGCCTCCTCGCCTCCCAAAGTG 120
           ||||||||||||||||||||||||| |||||||||||| |||||| || ||||||||| |
Db  9820 GCCAGACTGGTCTCGAACTCCTGACCTCAGGTGATCTGCCCGCCTCGGCCTCCCAAAGTG 9761

Qy   121 CTGGGATTATAGGGGTGAGCGCCGCCACCGCGCCCAGCCAACATTTTTAAATACTGAAAGTAG 180
           |||||||||||| |||||||||| |||||| || |||||||||||||| |||||||||| ||
Db  9760 CTGGGATTACAGGCGTGACCACCGTACCCGGCCTTTTGTTGGTCTTAATTTATACTTT 9701

Qy   181 AGGGAATAGTTATAGTGTACCCCATTTACCCATCACTCAGTTCAAACAGCCTGGTGACATA 240
           |||||| |||||| ||||||| || |||| ||| |||||||| ||||||||||| ||||
Db  9700 CTTGATGAATATTGGGTGTTAAGCATTT---TTTCAAGTACTTATGAGCCATTGTTATATC 9644

Qy   241 TTTATTTCTTCTATACCAGTACCGTACTCTCCCCACTGGGATTATTTTTAAGGCAAAACCC 300
           || || || ||||| ||||| |||||||||||||||||| ||||| |||||| || |||
Db  9643 TTCTTTCGTGGAATCTCTGTTTAAATATTTTGCTCATTTAAAAAAAATTTTAGGCTGGGCAT 9584

Qy   301 AGAIG-----ACATTTTATCCCTAAATACTT--TAGAIAAAGGTGTTCTTTTGGAAAAAAA 352
           |||      |||||||| ||||||||||| |  ||| ||||||||| |||||||||||||
Db  9583 GGTGGCGTCACACCTGTAATCCTAGCATTTTTGGGAGGCCGAGGTGGGATTGCTTGAGG 9524

Qy   353 TCATAACCTCAGGACCAGCCAGCCTGGGCCAACATGGTGAAACCCTGTCTACTAAAAATACA 412
           || | || || || || ||||||||||| |||| | | |||||| | || || ||||| ||
Db  9523 TCAGGAGTTTCAGACCAGCCTGGCCAACATAGTGAAACCTCATCTCTACTAAAAATACAA 9464

Qy   413 AAATTAGCCTTGGCATGGTGGTGGGCGCCTGTAATCCCAGCTACTCAGGGAGGCTGAGGCAG 472
           ||||||||| ||||| ||| || |||||| ||||||||||| ||| ||||||| ||| |||
Db  9463 AATTTAGCTGGGCCGTGGTGGCGGGCGCCTGTAATCCCAGCTACTCGGGGAGGCTGAGGCAG 9404

Qy   473 GAGAATCACTTGAACCCGGGAAGCAGAGGTTGCAGTGAGCTGAGATCGTGCCATTGCACT 532
           |||| || || || |||||||| || | ||||||||| ||||| ||| || || ||||
Db  9403 GAGGATCGCGTTGAACCCAGGAGGCGGAGGTTGCAGTGAGCCAAGATCATACTACTGCACT 9344

Qy   533 GGGCAGCAGAGACAGA-AATGAAACTCTGTCTCAAAAACAAACAAACAAAAAACCACTAT 591
           |||| || ||| | ||  | |||| | || || |||||||||||||| |||||| |||
Db  9343 CCAGCCTGGGCAACAAGAGTGAAAACTTCGTCTCAAAAAAAAAAAAAAAAAGAAAAT 9284
```

RESULT 10

```
PCT-US01-01354-24569/c
Sequence 24569, Application PC/TUS0101354
GENERAL INFORMATION:
   APPLICANT: Human Genome Sciences, Inc. et al.
   TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
   FILE REFERENCE: PC004PCT
   CURRENT APPLICATION NUMBER: PCT/US01/01354
   CURRENT FILING DATE: 2001-03-17
   NUMBER OF SEQ ID NOS: 42506
   SOFTWARE: Patentin Ver. 2.0
SEQ ID NO 24569
   LENGTH: 10702
   TYPE: DNA
   ORGANISM: Homo sapiens
PCT-US01-01354-24569

Query Match           22.6%;  Score 226;  DB 1;  Length 10702;
Best Local Similarity 64.48%;  Pred. No. 1.1e-20;
Matches 390;  Conservative 0;  Mismatches 205;  Indels 11;  Gaps 3;

Qy   4  GCCACCAATGCCTGGCCTAATTTCTTATTTTTAGTAGAGACGAGGTTTTGCCATGTTGACC  63
        |||||| || | ||||  | ||||||||||| ||  ||||| || ||| | ||||||| ||
Db 910  GTCACCAAGCCCAGCTAGTTTTTTGTTATTTTTAGCAAAGACGGGTTTCACCATGTTGGCC 851

Qy  64  AGGCTGGTCTCGAAACTCTTGACCTCAGGTGATCCGCCTGCCTCAGCCTCCCAAAGTGCTG 123
        |||||||||| ||||||| ||||||||||||||||| ||||||||  ||||||||||||||
Db 850  AGGCTGGTCTTGAAACTCCTGACCTCAGGTGATCCACCTGCCTCGGCCTCCCAAAGTGCTG 791

Qy 124  GGATTATAGGCGTGAGCGCCGCGCCCACCCAGCCACCAACATTTTTAAATACTGAAAGTAGAGG 183
        |||||| |||||| ||||||  | |||  |  || |      |||||   | || |   |||
Db 790  GGATTACAGGCGTTGAGCGCCACCGCGCCCGGCCTTTTTATGTCTGTATTCT 731

Qy 184  GAATAGTTATATGTGTACCCCATTTACCCATCACTCAGTTCAACAGCTGGTGACATATTT 243
        |    || |  |       |||||  |||| | ||||  |  ||| | |  ||| ||| ||
Db 730  GTGCATATTCTTCTCACGGATAATAATAAAACTTCCAGACACTGGGCTT 671

Qy 244  ATTTCTTCTATACCAGTACCGTACTCTCCCCACTGGGATTATTTTAGGCAAAAC--CA 301
        |  | ||| | || | ||    | ||   ||  | |  | |||||| |||||||  ||
Db 670  AATAGTCTTAAAAATACTTTAGATATATATATTTTTTTTTCAGAAATATTTCCAGTA 611

Qy 302  GATGACATTTTATCCCTTAAATACTTTGAGATAAAGGTGTTCTTTGAAAAAAATCATAACCT 361
        || | || ||  | |  |||  ||||  |||  || | | ||   | || ||| |||| |||
Db 610  GAGAACTAGAAATCATAAAAGAAAAAAATTTAGCGACGGGATCACCTGAGGGTCAGGAGTT 551

Qy 362  CAGGACCAGCCTGGCCAACATGGTGAAACCCTGTCTCTGTACTAAAAATACAAAAATTAGCT 421
        ||| |||||||||||||  ||||| ||||||  |||||  ||||||||||||||||||||| 
Db 550  CAAGACCAGCCTGGCCACACATGGCGAAACCCCATCTCTACTAAAAATACAAAAATTAGCC 491

Qy 422  TGGGCATGGTGGTGGGCGCACCTGTAATCCCAGCTACTCAGGAGGCTGAGGCAGAATCAC 481
         ||||| |||||||||||||  ||||||||||  |||| | |||||||||||||| |||||
Db 490  GGGCGTTGGTGGTGGGCGCCTGTAATCCCAACTACTTGGGAGGCTGAGGCAGGAGAATCAC 431

Qy 482  TTGAATCCGGGAAGCAGAGATTGCAGTGAGCTGAGA---TTGCGAGTCGAGGCCTGGGCGGAC 538
        ||||| ||||  |   | ||||||||||||| |||||   || |  | | ||||||||||||
Db 430  TTGAACCCGGAGGGTTGCAGTGAGCCGAGATCTTTGCCACTCCAGCCTGGGCGGAC 371

Qy 539  AGAGACAGAAATGAAACTCTGTCTCAAAAACAAACAAACAAAAACCACTATACATAAA 598
        |   |   |||||||||  |  | |||||  ||||||||  ||||  ||  |    |||
Db 370  -----AAGAGTGAAACGCCATCTCAAAAAAAAAAAAAAAAAAAGCCAAAATAAAA 317

Qy 599  AATGAA 604
         || |
Db 316  ATTGGA 311
```

Priority data (right column):

CURRENT FILING DATE: 2001-01-17
PRIOR FILING DATE: 2000-01-31; PRIOR APPLICATION NUMBER: 60/179,065
PRIOR FILING DATE: 2000-02-04; PRIOR APPLICATION NUMBER: 60/180,628
PRIOR FILING DATE: 2000-06-28; PRIOR APPLICATION NUMBER: 60/216,886
PRIOR FILING DATE: 2000-07-07; PRIOR APPLICATION NUMBER: 60/214,886
PRIOR FILING DATE: 2000-07-07; PRIOR APPLICATION NUMBER: 60/216,647
PRIOR FILING DATE: 2000-08-14; PRIOR APPLICATION NUMBER: 60/225,267
PRIOR FILING DATE: 2000-07-07; PRIOR APPLICATION NUMBER: 60/216,880
PRIOR FILING DATE: 2000-07-11; PRIOR APPLICATION NUMBER: 60/217,487
PRIOR FILING DATE: 2000-08-14; PRIOR APPLICATION NUMBER: 60/225,447
PRIOR FILING DATE: 2000-08-14; PRIOR APPLICATION NUMBER: 60/225,758
PRIOR FILING DATE: 2000-07-11; PRIOR APPLICATION NUMBER: 60/217,496
PRIOR FILING DATE: 2000-07-26; PRIOR APPLICATION NUMBER: 60/220,963
PRIOR FILING DATE: 2000-08-14; PRIOR APPLICATION NUMBER: 60/225,270
PRIOR FILING DATE: 2000-07-11; PRIOR APPLICATION NUMBER: 60/216,880
PRIOR FILING DATE: 2000-08-14; PRIOR APPLICATION NUMBER: 60/225,834
PRIOR FILING DATE: 2000-12-08; PRIOR APPLICATION NUMBER: 60/251,869
PRIOR FILING DATE: 2000-09-27; PRIOR APPLICATION NUMBER: 60/235,834
PRIOR FILING DATE: 2000-09-21; PRIOR APPLICATION NUMBER: 60/234,223
PRIOR FILING DATE: 2000-09-21; PRIOR APPLICATION NUMBER: 60/234,274
PRIOR FILING DATE: 2000-08-30; PRIOR APPLICATION NUMBER: 60/228,924
PRIOR FILING DATE: 2000-08-14; PRIOR APPLICATION NUMBER: 60/224,518
PRIOR FILING DATE: 2000-09-29; PRIOR APPLICATION NUMBER: 60/236,369
PRIOR FILING DATE: 2000-07-26; PRIOR APPLICATION NUMBER: 60/220,964
PRIOR FILING DATE: 2000-08-14; PRIOR APPLICATION NUMBER: 60/224,519
PRIOR FILING DATE: 2000-09-29; PRIOR APPLICATION NUMBER: 60/236,369
PRIOR FILING DATE: 2000-11-17; PRIOR APPLICATION NUMBER: 60/249,299
PRIOR FILING DATE: 2000-10-20; PRIOR APPLICATION NUMBER: 60/241,809
PRIOR FILING DATE: 2000-10-20; PRIOR APPLICATION NUMBER: 60/241,809
PRIOR FILING DATE: 2000-11-01; PRIOR APPLICATION NUMBER: 60/244,617
PRIOR FILING DATE: 2000-10-20; PRIOR APPLICATION NUMBER: 60/241,785
PRIOR FILING DATE: 2000-09-29; PRIOR APPLICATION NUMBER: 60/236,327
PRIOR FILING DATE: 2000-09-29; PRIOR APPLICATION NUMBER: 60/236,368
PRIOR FILING DATE: 2000-08-14; PRIOR APPLICATION NUMBER: 60/225,268
PRIOR FILING DATE: 2000-12-08; PRIOR APPLICATION NUMBER: 60/251,868
PRIOR FILING DATE: 2000-12-08; PRIOR APPLICATION NUMBER: 60/251,856
PRIOR FILING DATE: 2000-09-01; PRIOR APPLICATION NUMBER: 60/229,344
PRIOR FILING DATE: 2000-12-08; PRIOR APPLICATION NUMBER: 60/234,997
PRIOR FILING DATE: 2000-09-25; PRIOR APPLICATION NUMBER: 60/234,997
PRIOR FILING DATE: 2000-09-01; PRIOR APPLICATION NUMBER: 60/229,343

```
RESULT 11
US-09-764-905-24569/c
Sequence 24569, Application US/09764905
GENERAL INFORMATION:
   APPLICANT: Rosen et al.
   TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
   FILE REFERENCE: PC004
   CURRENT APPLICATION NUMBER: US/09/764,905
```

PRIOR APPLICATION NUMBER: 60/229,345
PRIOR FILING DATE: 2000-09-01
PRIOR APPLICATION NUMBER: 60/229,287
PRIOR FILING DATE: 2000-09-01
PRIOR APPLICATION NUMBER: 60/229,513
PRIOR FILING DATE: 2000-09-05
PRIOR APPLICATION NUMBER: 60/231,413
PRIOR FILING DATE: 2000-09-08
PRIOR APPLICATION NUMBER: 60/229,509
PRIOR FILING DATE: 2000-09-05
PRIOR APPLICATION NUMBER: 60/236,367
PRIOR FILING DATE: 2000-11-17
PRIOR APPLICATION NUMBER: 60/237,039
PRIOR FILING DATE: 2000-09-05
PRIOR APPLICATION NUMBER: 60/237,040
PRIOR FILING DATE: 2000-10-02
PRIOR APPLICATION NUMBER: 60/237,037
PRIOR FILING DATE: 2000-10-02
PRIOR APPLICATION NUMBER: 60/236,802
PRIOR FILING DATE: 2000-10-02
PRIOR APPLICATION NUMBER: 60/236,370
PRIOR FILING DATE: 2000-10-02
PRIOR APPLICATION NUMBER: 60/237,038
PRIOR FILING DATE: 2000-10-02
PRIOR APPLICATION NUMBER: 60/239,935
PRIOR FILING DATE: 2000-10-13
PRIOR APPLICATION NUMBER: 60/239,937
PRIOR FILING DATE: 2000-10-13
PRIOR APPLICATION NUMBER: 60/241,787
PRIOR FILING DATE: 2000-10-20
PRIOR APPLICATION NUMBER: 60/246,474
PRIOR FILING DATE: 2000-11-08
PRIOR APPLICATION NUMBER: 60/246,532
PRIOR FILING DATE: 2000-11-08
PRIOR APPLICATION NUMBER: 60/249,216
PRIOR FILING DATE: 2000-11-17
PRIOR APPLICATION NUMBER: 60/249,210
PRIOR FILING DATE: 2000-11-17
PRIOR APPLICATION NUMBER: 60/226,681
PRIOR FILING DATE: 2000-08-22
PRIOR APPLICATION NUMBER: 60/225,759
PRIOR FILING DATE: 2000-08-14
PRIOR APPLICATION NUMBER: 60/225,266
PRIOR FILING DATE: 2000-08-14
PRIOR APPLICATION NUMBER: 60/225,213
PRIOR FILING DATE: 2000-08-14
PRIOR APPLICATION NUMBER: 60/225,213
PRIOR FILING DATE: 2000-08-14
PRIOR APPLICATION NUMBER: 60/227,182
PRIOR FILING DATE: 2000-08-22
PRIOR APPLICATION NUMBER: 60/225,214
PRIOR FILING DATE: 2000-08-14
PRIOR APPLICATION NUMBER: 60/235,836
PRIOR FILING DATE: 2000-09-27
PRIOR APPLICATION NUMBER: 60/230,438
PRIOR FILING DATE: 2000-09-06
PRIOR APPLICATION NUMBER: 60/215,135
PRIOR FILING DATE: 2000-06-30

PRIOR FILING DATE: 2000-11-17
PRIOR APPLICATION NUMBER: 60/249,211
PRIOR FILING DATE: 2000-11-17
PRIOR APPLICATION NUMBER: 60/249,215
PRIOR FILING DATE: 2000-11-17
PRIOR APPLICATION NUMBER: 60/249,264
PRIOR FILING DATE: 2000-11-17
PRIOR APPLICATION NUMBER: 60/249,214
PRIOR FILING DATE: 2000-11-17
PRIOR APPLICATION NUMBER: 60/249,297
PRIOR FILING DATE: 2000-11-17
PRIOR APPLICATION NUMBER: 60/249,400
PRIOR FILING DATE: 2000-09-14
PRIOR APPLICATION NUMBER: 60/231,242
PRIOR FILING DATE: 2000-09-08
PRIOR APPLICATION NUMBER: 60/232,081
PRIOR FILING DATE: 2000-09-08
PRIOR APPLICATION NUMBER: 60/232,080
PRIOR FILING DATE: 2000-09-08
PRIOR APPLICATION NUMBER: 60/231,414
PRIOR FILING DATE: 2000-09-08
PRIOR APPLICATION NUMBER: 60/231,244
PRIOR FILING DATE: 2000-09-08
PRIOR APPLICATION NUMBER: 60/233,063
PRIOR FILING DATE: 2000-09-14
PRIOR APPLICATION NUMBER: 60/232,397
PRIOR FILING DATE: 2000-09-14
PRIOR APPLICATION NUMBER: 60/232,399
PRIOR FILING DATE: 2000-09-14
PRIOR APPLICATION NUMBER: 60/232,401
PRIOR FILING DATE: 2000-09-14
PRIOR APPLICATION NUMBER: 60/241,808
PRIOR FILING DATE: 2000-10-20
PRIOR APPLICATION NUMBER: 60/241,826
PRIOR FILING DATE: 2000-10-20
PRIOR APPLICATION NUMBER: 60/241,786
PRIOR FILING DATE: 2000-10-20
PRIOR APPLICATION NUMBER: 60/241,221
PRIOR FILING DATE: 2000-10-20
PRIOR APPLICATION NUMBER: 60/246,475
PRIOR FILING DATE: 2000-11-08
PRIOR APPLICATION NUMBER: 60/231,243
PRIOR FILING DATE: 2000-09-08
PRIOR APPLICATION NUMBER: 60/233,065
PRIOR FILING DATE: 2000-09-14
PRIOR APPLICATION NUMBER: 60/232,398

Query Match
Best Local Similarity   22.6%;   Score 226;   DB 30;   Length 10702;
Matches 390; Conservative 64.4%; Pred. No. 1.1e-20; Indels 11; Gaps 3;
                         0; Mismatches 205;

```
Qy    4 GCCACCATGCCTGGCGTAATTTTCTTATTTTTAGTAGAGACGAGGTTTTGCCATGTTGACC   63
        |||||| ||| |||| |||||||||||| |||||||| |||||||||| ||||||||| ||
Db  910 GTCACCAAGCCCAGCTAGTTTTGTAATTTTTAGCAAAGACGGGGTTTCACCATGTTGGCC   851

Qy   64 AGGCTGGGTCTCGAACTCTTGACCTCAGGTGATCCGCCCGCCTCGCCTCCCAAAGTGCTG   123
        |||||||||| ||||| ||||| |||||||| ||||||||| ||||| |||||||||||||
Db  850 AGGCTGGGTCTTGAACTCCCTGACCTCAGGTGATCCACCTGCCTCAGCCTCCCAAAGTGCTG 791

Qy  124 GGATTATAGGCGTGAGCCGCGCGCACCCAGCCGCAACATTTTTAAATACTGAAAAGTAGAGG 183
        ||||| | |||| ||||||||||| ||| ||| ||| |      |  |      |
Db  790 GGATTACAGGCCTTGAGCCACCGCGCCCGGCGCTGCGCTCGCCTATATTTTAAGTCTGTATTCT 731

Qy  184 GAATAGTTTATAGTGTTACCCCATTTACCCATCACTCAGTTTCAACAGCTGGTGACATATTT  243
        |||   | |   |  || |    |
Db  730 GTGCATTATTCTTCTCAGCAGATAAATATAAAACTTCCACAGGGGCTTCCAGACACTGGGCTT 671

Qy  244 ATTTCTTCTATACCAGTACCGTACTCTCCCCACTGGGGATTATTTTTTTTAAGGCAAAACC--CA 301
        ||  |   |   |       |       |
Db  670 AATAGTCTTAAAAATATATATATATATATATATATATTTTTTTTTTTCAGAAATATTTCCAGTA 611
```

```
Qy   302  GATGACATTTTATCCCTAAAATACTTTAGATAATAAAGGTGTTCTTTGAAAAAATCATATAACCT  361
               |||    ||
Db   610  GAGAACTAGAAATCATAAAAGAAAAAATTTAGCGAAGCGGGATCACCTGAGGTCAGGAGTT  551

Qy   362  CAGGACCAGCCTGGCCAACATGGTGAAAACCCTGTCTCTACTAAAAATACAAAAATTAGCT  421
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   316  ATTGGA  311
```

RESULT 12
US-10-092-399-24569/c
; Sequence 24569, Application US/10092399
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
; FILE REFERENCE: PC04C1
; CURRENT APPLICATION NUMBER: US/10/092,399
; CURRENT FILING DATE: 2002-03-07
; NUMBER OF SEQ ID NOS: 42506
; Prior Application removed - See File Wrapper or Palm
; SEQ ID NO 24569
; LENGTH: 10702
; TYPE: DNA
; ORGANISM: Homo sapiens
; SOFTWARE: PatentIn Ver. 2.0
US-10-092-399-24569

Query Match   22.6%;
Best Local Similarity   64.48%;
Matches 390; Conservative 0; Mismatches 205; Indels 11; Gaps 3;
              Score 226; DB 39; Length 10702; Pred. No. 1.1e-20;

```
Qy     4  GCCACCATGCCTGGCCTAAATTTCTTATTTTATTTTTPAGTAGAGACGAGGTTTGCCATGTTGACC  63
          |||  |  ||    ||
Db   910  GTCACCAAGCCCAGCCTAGTTTTTGTTGTATTTTTAGCAAAGACGGGGTTTCACCATGTTGGCC  851

Qy    64  AGGCTGGTCTCGAACTCTTGACCTCAGGTGATCCGCCTGCCTCAGCCTCCCAAAGTGCTG  123
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   850  AGGCTGGTCTTGAACTCCTGACCTCAGGTGATCCACCTGCCTCCCAAAGTGCTG  791

Qy   124  GGATTATAGGCGTGAGCGCGCCGCCACCGCCAACACATTTTTTTAAATACTGAAAAGTAGAGG  183
          |||||||||||| |||| |||| ||    ||      |||||   |||
Db   790  GGATTACAGGCGTTGAGCCACCACCGCCCGCCTGCCTCTGCATATTTTTATGTCTGTATTCT  731

Qy   184  GAATAGTTATAGTGTACCCCATTTACCCTACACTCAGTTTCAACACAGCTGGTGACATATTT  243
          |||         || |  ||||  |  ||
Db   730  GTGCATATTCTTCTCCACGAGATAATATAAAAACTTCCACAGGGGGTCCAGACACTGGGGCTT  671

Qy   244  ATTTCTTCTATACCAGTACCGTACTCTCCCCACTGGGATTATTTTAAGGCAAAAACC--CA  301
          ||| ||||    |||     ||       |||
Db   670  AATAGTCTTAAAAATATATATATATATATAGATAAAGGTGTTCTTTGAAAAAATCATTAACCT  611

Qy   302  GATGACATTTTATCCCTAAAATACTTTGATAATAAAAGGTGTTCTTTGAAAAAATCATATAACCT  361
          ||||                |||
Db   670  AATAGTCTTAAAAATATATATATATATATAGATAAAGGTGTTCTTTGAAAAAATCATTAACCT  611

Qy   362  CAGGACCAGCCTGGCCAACATGGTGAAACCCTGTCTCTACTAAAAATACAAAAATTAGCT  421
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   610  GAGAACTAGAAATCATAAAAGAAAAAATTTAGCGGAGCGGGATCACCTGAGGTCAGGAGTT  551
```

RESULT 13
US-09-973-278-780/c
; Sequence 780, Application US/09973278
; GENERAL INFORMATION:
; APPLICANT: Fischer et al.
; TITLE OF INVENTION: 123 Human Secreted Proteins
; FILE REFERENCE: P2010P2
; CURRENT APPLICATION NUMBER: US/09/973,278
; CURRENT FILING DATE: 2001-10-10
; PRIOR APPLICATION NUMBER: 60/239,899
; PRIOR FILING DATE: 2000-10-13
; PRIOR APPLICATION NUMBER: 09/227,357
; PRIOR FILING DATE: 1999-01-08
; PRIOR APPLICATION NUMBER: PCT/US98/13684
; PRIOR FILING DATE: 1998-07-07
; PRIOR APPLICATION NUMBER: 60/051,926
; PRIOR FILING DATE: 1997-07-08
; PRIOR APPLICATION NUMBER: 60/052,793
; PRIOR FILING DATE: 1997-07-08
; PRIOR APPLICATION NUMBER: 60/051,925
; PRIOR FILING DATE: 1997-07-08
; PRIOR APPLICATION NUMBER: 60/051,929
; PRIOR FILING DATE: 1997-07-08
; PRIOR APPLICATION NUMBER: 60/052,803
; PRIOR FILING DATE: 1997-07-08
; PRIOR APPLICATION NUMBER: 60/052,732
; PRIOR FILING DATE: 1997-07-08
; PRIOR APPLICATION NUMBER: 60/051,916
; PRIOR FILING DATE: 1997-07-08
; PRIOR APPLICATION NUMBER: 60/051,932
; PRIOR FILING DATE: 1997-07-08
; PRIOR APPLICATION NUMBER: 60/051,930
; PRIOR FILING DATE: 1997-07-08
; PRIOR APPLICATION NUMBER: 60/051,918
; PRIOR FILING DATE: 1997-07-08
; PRIOR APPLICATION NUMBER: 60/051,920
; PRIOR FILING DATE: 1997-07-08
; PRIOR APPLICATION NUMBER: 60/051,919
; PRIOR FILING DATE: 1997-07-08
; PRIOR APPLICATION NUMBER: 60/052,795
; PRIOR FILING DATE: 1997-07-08
; PRIOR APPLICATION NUMBER: 60/052,733
; PRIOR FILING DATE: 1997-07-08
; PRIOR APPLICATION NUMBER: 60/051,931
; PRIOR FILING DATE: 1997-07-08
; PRIOR APPLICATION NUMBER: 60/051,928
; PRIOR FILING DATE: 1997-07-08
; PRIOR APPLICATION NUMBER: 60/051,928
; PRIOR FILING DATE: 1997-07-08
; PRIOR APPLICATION NUMBER: 60/055,722
; PRIOR FILING DATE: 1997-08-18
; PRIOR APPLICATION NUMBER: 60/055,723
; PRIOR FILING DATE: 1997-08-18
; PRIOR APPLICATION NUMBER: 60/055,948
; PRIOR FILING DATE: 1997-08-18
; PRIOR APPLICATION NUMBER: 60/055,949

```
;  PRIOR FILING DATE: 1997-08-18
;  PRIOR APPLICATION NUMBER: 60/055,953
;  PRIOR FILING DATE: 1997-08-18
;  PRIOR APPLICATION NUMBER: 60/055,950
;  PRIOR FILING DATE: 1997-08-18
;  PRIOR APPLICATION NUMBER: 60/055,947
;  PRIOR FILING DATE: 1997-08-18
;  PRIOR APPLICATION NUMBER: 60/055,964
;  PRIOR FILING DATE: 1997-08-18
;  PRIOR APPLICATION NUMBER: 60/055,964
;  PRIOR FILING DATE: 1997-08-18
;  PRIOR APPLICATION NUMBER: 60/056,360
;  PRIOR FILING DATE: 1997-08-18
;  PRIOR APPLICATION NUMBER: 60/055,684
;  PRIOR FILING DATE: 1997-08-18
;  PRIOR APPLICATION NUMBER: 60/055,984
;  PRIOR FILING DATE: 1997-08-18
;  PRIOR APPLICATION NUMBER: 60/055,954
;  PRIOR FILING DATE: 1997-09-12
;  PRIOR APPLICATION NUMBER: 60/058,785
;  PRIOR FILING DATE: 1997-09-12
;  PRIOR APPLICATION NUMBER: 60/058,664
;  PRIOR FILING DATE: 1997-09-12
;  PRIOR APPLICATION NUMBER: 60/058,660
;  PRIOR FILING DATE: 1997-09-12
;  PRIOR APPLICATION NUMBER: 60/058,661
;  PRIOR FILING DATE: 1997-09-12

;  NUMBER OF SEQ ID NOS: 947
;  SOFTWARE: Patentin Ver. 2.0

US-09-973-278-780
;  SEQ ID NO 780
;    LENGTH: 1503
;    TYPE: DNA
;    ORGANISM: Homo sapiens

Query Match          22.2%;  Score 222.4;  DB 36;  Length 1503;
Best Local Similarity 63.6%;  Pred. No. 4.7e-20;
Matches 392;  Conservative 0;  Mismatches 211;  Indels 13;  Gaps 3;

Qy     3 TGCCACCATGCCCTGGCTAATTTTTCTTATTTTTTTAGTAGAGAGACGAGGTTTTGCCANGTGAC    62
          ||||||||  |||||||||||||                                |||| |||||| |
Db  1357 TGCCACCACCTGGCTAATTTTTGTATTTTTAGTAGAGACTGGGGTTTCACCATGTTGGC       1298

Qy    63 CAGGCTGGTCTCGAACTCTTGACCTCAGGTGATCCGCGCCTTGCCTCCCAAAGTGCT   122
          ||||| | ||||||||| ||||||||||||||||| |  || || |||| ||||||||
Db  1297 CAGGCTAGTCTCGAACTCCTGACCTGATCCGCCGCTTCGGCCTTCCCAAAGTGCT  1238

Qy   123 GGGATTATAGGCGTGAGCGCGCCGCACCCAGCCAACACATTTTTAAATACTGAAAAGTAGAG   182
          |||| ||||||| |||||  |||||||  |||||  ||| ||||| ||| ||||  ||   |
Db  1237 GGGAGTATAGGCATGAGCCACCACAGGCCCAGCCTTATAAATGTGCATATC---CATGTGTGT  1181

Qy   183 GGAATAGTTATAGTGTGTACCCCCATTTACCCATCACTCAGTTTCAAACAGCTGGTGACATATT   242
          |  |||                   || |||  ||| |||
Db  1180 ATATACTGTTACATACATCATCATCTCCCACTGGGATTATATTTAAGGCAAAACCCAG   1121

Qy   243 TATTTCTTTCTATACCACTGGCTCTCCCACTGGGATTATATTTAAGGCAAAACCCAG   302
          || |  | | ||    | |  ||
Db  1120 TAATACGCCAATATGTCAAGGAAATTGTAAAGGACACAGTGGCTC   1061

Qy   303 ATGACATTTTATCCCTAAATACTTTAGATAAAAGGT---GTTCTTTGAAAAAATCATAACC   360
          |||     ||| || |   | |  | || |   ||| |
Db  1060 ATGCCTCTAATCCGAACACTCTGGGAGCCTGAGGCGGGTGGATCACCTGAGGTCAGGAGT   1001

Qy   361 TCAGGACCAGCCTGGCCAAACATGGTGAAACCCTGTCTGTACTAAAAATACAAAAATTAGC   420
          || || |||||||||||||| ||||||||| | |||||  ||||||||||||||||||||||
Db  1000 TCGAGGCCAGCCTGGCCAACATGGTGAAACACTGTCTCTACTAAAAATACAAAAATTAGC   941

Qy   421 TTGGGCATGGTGGTGGGCACCTGTAATCCCAGCTACTCAGGAGGCTGAGGCAGGAGAATCA   480
          | ||||||||||| |  |||| ||||||| |||||   |||||||| |||||||||||||
Db   940 CAGGCATGGTGGTGTACCCCCTCTGTAATCCCCAGCTATTCGGGAGGCCGAGGCAGGAGAATCG   881

Qy   481 CTTGAATCCGGGAGGCAGGAGATTGCAGTGAGCTGAGAT-------TGCAGTCGCGAGCCTG   532
          ||||||  | |||||||| ||| |||||||||| |||| |       |||||| | |||||||
Db   880 CTTGAACCTGGGAGGCGGAGGTTGCAGTGAGCCGAGAAATTCAACCACTCCACTCCAGCCTG   821

Qy   533 GGCGACAGAGCAGAAATGAAACTCTGTCTCAAAACAAACAAAACAAAAAAAAACCACTATA   592
          |||| |||||| |||| |||| | || || ||||| || |  ||||||  ||| || ||| |
Db   820 TGCAACAGAGAAAGACTCCATCTCAAAACACAAAACAAAAACAAAAAAAACCAAAAA   761

Qy   593 CATAAAAATGAACAAT   608
          ||  ||| |||||||
Db   760 AAAGGAAGAGATATAT   745
```

```
RESULT 14
US-09-950-083-9471
;  Sequence 9471, Application US/09950083
;  GENERAL INFORMATION:
;    APPLICANT: Rosen, et. al
;    TITLE OF INVENTION: Human Secreted Proteins
;    FILE REFERENCE: PS805
;    CURRENT APPLICATION NUMBER: US/09/950,083
;    CURRENT FILING DATE: 2001-09-12
;    PRIOR APPLICATION NUMBER: 60/278,650
;    PRIOR FILING DATE: 2001-03-27
;    PRIOR APPLICATION NUMBER: 09/833,245
;    PRIOR FILING DATE: 2001-04-12
;    PRIOR APPLICATION NUMBER: PCT/US01/11988
;    PRIOR FILING DATE: 2001-04-12
;    PRIOR APPLICATION NUMBER: PCT/US00/06824
;    PRIOR FILING DATE: 2000-03-09
;    PRIOR APPLICATION NUMBER: PCT/US00/06057
;    PRIOR FILING DATE: 2000-03-09
;    PRIOR APPLICATION NUMBER: PCT/US00/06049
;    PRIOR FILING DATE: 2000-03-09
;    PRIOR APPLICATION NUMBER: PCT/US00/06013
;    PRIOR FILING DATE: 2000-03-09
;    PRIOR APPLICATION NUMBER: PCT/US00/06014
;    PRIOR FILING DATE: 2000-03-09
;    PRIOR APPLICATION NUMBER: PCT/US00/06042
;    PRIOR FILING DATE: 2000-03-09
;    PRIOR APPLICATION NUMBER: PCT/US00/06059
;    PRIOR FILING DATE: 2000-03-09
;    PRIOR APPLICATION NUMBER: PCT/US00/06058
;    PRIOR FILING DATE: 2000-03-09
;    PRIOR APPLICATION NUMBER: PCT/US00/06044
;    PRIOR FILING DATE: 2000-03-09
;    PRIOR APPLICATION NUMBER: PCT/US00/06012
;    PRIOR FILING DATE: 2000-03-09
;    PRIOR APPLICATION NUMBER: PCT/US00/06043
;    PRIOR FILING DATE: 2000-03-09
;    PRIOR APPLICATION NUMBER: PCT/US00/06765
;    PRIOR FILING DATE: 2000-03-09
;    PRIOR APPLICATION NUMBER: PCT/US00/06792
;    PRIOR FILING DATE: 2000-03-16
;    PRIOR APPLICATION NUMBER: PCT/US00/06791
;    PRIOR FILING DATE: 2000-03-16
;    PRIOR APPLICATION NUMBER: PCT/US00/06822
;    PRIOR FILING DATE: 2000-03-16
;    PRIOR APPLICATION NUMBER: PCT/US00/06782
;    PRIOR FILING DATE: 2000-03-16
;    PRIOR APPLICATION NUMBER: PCT/US00/06830
;    PRIOR FILING DATE: 2000-03-16
;    PRIOR APPLICATION NUMBER: PCT/US00/06823
;    PRIOR FILING DATE: 2000-03-16
;    PRIOR APPLICATION NUMBER: PCT/US00/06828
;    PRIOR FILING DATE: 2000-03-16
;    PRIOR APPLICATION NUMBER: PCT/US00/06781
;    PRIOR FILING DATE: 2000-03-16
;    PRIOR APPLICATION NUMBER: PCT/US00/07505
;    PRIOR FILING DATE: 2000-03-22
;    PRIOR APPLICATION NUMBER: PCT/US00/07440
;    PRIOR FILING DATE: 2000-03-22
;    PRIOR APPLICATION NUMBER: PCT/US00/07506
;    PRIOR FILING DATE: 2000-03-22
;    PRIOR APPLICATION NUMBER: PCT/US00/07507
;    PRIOR FILING DATE: 2000-03-22
;    PRIOR APPLICATION NUMBER: PCT/US00/07535
```

PRIOR FILING DATE: 2000-03-22
PRIOR APPLICATION NUMBER: PCT/US00/07525
PRIOR FILING DATE: 2000-03-22
PRIOR APPLICATION NUMBER: PCT/US00/07527
PRIOR FILING DATE: 2000-03-22
PRIOR APPLICATION NUMBER: PCT/US00/07661
PRIOR FILING DATE: 2000-03-22
PRIOR APPLICATION NUMBER: PCT/US00/07534
PRIOR FILING DATE: 2000-03-22
PRIOR APPLICATION NUMBER: PCT/US00/07483
PRIOR FILING DATE: 2000-03-22
PRIOR APPLICATION NUMBER: PCT/US00/07526
PRIOR FILING DATE: 2000-06-01
PRIOR APPLICATION NUMBER: PCT/US00/14929
PRIOR FILING DATE: 2000-03-23
PRIOR APPLICATION NUMBER: PCT/US00/07724
PRIOR FILING DATE: 2000-03-23
PRIOR APPLICATION NUMBER: PCT/US00/07723
PRIOR FILING DATE: 2000-03-23
PRIOR APPLICATION NUMBER: PCT/US00/07726
PRIOR FILING DATE: 2000-03-23
PRIOR APPLICATION NUMBER: PCT/US00/07578
PRIOR FILING DATE: 2000-03-23
PRIOR APPLICATION NUMBER: PCT/US00/07722
PRIOR FILING DATE: 2000-03-23
PRIOR APPLICATION NUMBER: PCT/US00/07677
PRIOR FILING DATE: 2000-03-23
PRIOR APPLICATION NUMBER: PCT/US00/07725
PRIOR FILING DATE: 2000-04-06
PRIOR APPLICATION NUMBER: PCT/US00/08983
PRIOR FILING DATE: 2000-04-06
PRIOR APPLICATION NUMBER: PCT/US00/08982
PRIOR FILING DATE: 2000-04-06
PRIOR APPLICATION NUMBER: PCT/US00/09070
PRIOR FILING DATE: 2000-04-06
PRIOR APPLICATION NUMBER: PCT/US00/09066
PRIOR FILING DATE: 2000-04-06
PRIOR APPLICATION NUMBER: PCT/US00/09067
PRIOR FILING DATE: 2000-04-06
PRIOR APPLICATION NUMBER: PCT/US00/09068
PRIOR FILING DATE: 2000-04-06
PRIOR APPLICATION NUMBER: PCT/US00/09069
PRIOR FILING DATE: 2000-04-06
PRIOR APPLICATION NUMBER: PCT/US00/08981
PRIOR FILING DATE: 2000-04-06
PRIOR APPLICATION NUMBER: PCT/US00/08980
PRIOR FILING DATE: 2000-04-06
PRIOR APPLICATION NUMBER: PCT/US00/09071
PRIOR FILING DATE: 2000-04-06
PRIOR APPLICATION NUMBER: PCT/US00/15136
PRIOR FILING DATE: 2000-06-01
PRIOR APPLICATION NUMBER: PCT/US00/15135
PRIOR FILING DATE: 2000-06-01
PRIOR APPLICATION NUMBER: PCT/US00/14926
PRIOR FILING DATE: 2000-06-01
PRIOR APPLICATION NUMBER: PCT/US00/14963.
PRIOR FILING DATE: 2000-06-01
PRIOR APPLICATION NUMBER: PCT/US00/15137
PRIOR FILING DATE: 2000-06-01
PRIOR APPLICATION NUMBER: PCT/US00/14933
PRIOR FILING DATE: 2000-06-01
PRIOR APPLICATION NUMBER: PCT/US00/14934
PRIOR FILING DATE: 2000-06-01
PRIOR APPLICATION NUMBER: PCT/US00/15135
PRIOR FILING DATE: 2000-06-01
PRIOR APPLICATION NUMBER: PCT/US00/14928
PRIOR FILING DATE: 2000-06-01
PRIOR APPLICATION NUMBER: PCT/US00/14973
PRIOR FILING DATE: 2000-06-01
PRIOR APPLICATION NUMBER: PCT/US00/14964
PRIOR FILING DATE: 2000-06-01
PRIOR APPLICATION NUMBER: PCT/US00/15137
PRIOR FILING DATE: 2000-09-26
PRIOR APPLICATION NUMBER: PCT/US00/26376

PRIOR APPLICATION NUMBER: PCT/US00/26371
PRIOR FILING DATE: 2000-09-26
PRIOR APPLICATION NUMBER: PCT/US00/26324
PRIOR FILING DATE: 2000-09-26
PRIOR APPLICATION NUMBER: PCT/US00/26323
PRIOR FILING DATE: 2000-09-26
PRIOR APPLICATION NUMBER: PCT/US00/26337
PRIOR FILING DATE: 2000-09-26
PRIOR APPLICATION NUMBER: PCT/US00/26337
PRIOR FILING DATE: 2000-09-26
PRIOR APPLICATION NUMBER: PCT/US01/13318
PRIOR FILING DATE: 2001-04-27
PRIOR APPLICATION NUMBER: US 60/124,146
PRIOR FILING DATE: 1999-03-12
PRIOR APPLICATION NUMBER: US 60/167,061
PRIOR FILING DATE: 1999-11-23
PRIOR APPLICATION NUMBER: US 60/124,093
PRIOR FILING DATE: 1999-03-12
PRIOR APPLICATION NUMBER: US 60/166,989
PRIOR FILING DATE: 1999-11-23
PRIOR APPLICATION NUMBER: US 60/124,145
PRIOR FILING DATE: 1999-03-12
PRIOR APPLICATION NUMBER: US 60/124,099
PRIOR FILING DATE: 1999-12-03
PRIOR APPLICATION NUMBER: US 60/168,654
PRIOR FILING DATE: 1999-12-03
PRIOR APPLICATION NUMBER: US 60/168,661
PRIOR FILING DATE: 1999-03-12
PRIOR APPLICATION NUMBER: US 60/124,096
PRIOR FILING DATE: 1999-12-03
PRIOR APPLICATION NUMBER: US 60/168,663
PRIOR FILING DATE: 1999-03-12
PRIOR APPLICATION NUMBER: US 60/124,095
PRIOR FILING DATE: 1999-12-03
PRIOR APPLICATION NUMBER: US 60/168,622
PRIOR FILING DATE: 1999-03-12
PRIOR APPLICATION NUMBER: US 60/124,143
PRIOR FILING DATE: 1999-12-03
PRIOR APPLICATION NUMBER: US 60/138,598
PRIOR FILING DATE: 1999-06-11
PRIOR APPLICATION NUMBER: US 60/168,665
PRIOR FILING DATE: 1999-12-03
PRIOR APPLICATION NUMBER: US 60/125,360
PRIOR FILING DATE: 1999-03-19
PRIOR APPLICATION NUMBER: US 60/124,095
PRIOR FILING DATE: 1999-03-12
PRIOR APPLICATION NUMBER: US 60/124,095
PRIOR FILING DATE: 1999-12-03
PRIOR APPLICATION NUMBER: US 60/138,626
PRIOR FILING DATE: 1999-06-11
PRIOR APPLICATION NUMBER: US 60/124,144
PRIOR FILING DATE: 1999-03-12
PRIOR APPLICATION NUMBER: US 60/168,662
PRIOR FILING DATE: 1999-12-03
PRIOR APPLICATION NUMBER: US 60/168,667
PRIOR FILING DATE: 1999-12-03
PRIOR APPLICATION NUMBER: US 60/138,574
PRIOR FILING DATE: 1999-06-11
PRIOR APPLICATION NUMBER: US 60/168,597
PRIOR FILING DATE: 1999-12-03
PRIOR APPLICATION NUMBER: US 60/138,664
PRIOR FILING DATE: 1999-06-11
PRIOR APPLICATION NUMBER: US 60/168,666
PRIOR FILING DATE: 1999-12-03
PRIOR APPLICATION NUMBER: US 60/126,051
PRIOR FILING DATE: 1999-03-23
PRIOR APPLICATION NUMBER: US 60/125,359
PRIOR FILING DATE: 1999-03-19
PRIOR APPLICATION NUMBER: US 60/124,142
PRIOR FILING DATE: 1999-03-12
PRIOR APPLICATION NUMBER: US 60/169,906

Query Match
Best Local Similarity   21.8%;   Score 217.6;   DB 36;   Length 1932;
                         63.2%;   Pred. No. 1.9e-19;
Matches 373; Conservative 0; Mismatches 204; Indels 13; Gaps 2;

```
Qy     1 CCTGCCACCATGCCTGGCTAATTTCTTTATTTTTAGTAGAGACGAGGTTTTGCCATGTTG   60
         || ||| |||| |||| || ||||||||| ||||||||| ||||||||||||||||||||
Db  1313 CCGGCCCCCATACCAGCTAATTTTATATTTTTTAGTAGAGAGATGGGGTTTTGCCATGTTG 1372
```

```
RESULT 15
US-10-105-299-11622
Sequence 11622, Application US/10105299
;  GENERAL INFORMATION:
;  APPLICANT: Rosen, et al
;  TITLE OF INVENTION: Human Secreted Proteins
;  FILE REFERENCE: PS950
;  CURRENT APPLICATION NUMBER: US/10/105,299
;  CURRENT FILING DATE: 2002-03-26
;  NUMBER OF SEQ ID NOS: 15197
;  Prior Application removed
;  Prior Application removed - See File Wrapper or Palm
;  SOFTWARE: Patentin Ver. 2.0
;  SEQ ID NO 11622
;  LENGTH: 1932
;  TYPE: DNA
;  ORGANISM: Homo sapiens
US-10-105-299-11622

Query Match      21.8%;     Score 217.6;    DB 40;    Length 1932;
Best Local Similarity   63.28%;
Matches  373;  Conservative   0;  Mismatches  204;  Indels  13;  Gaps  2;
```

```
Qy      1 CCTGCCACCATGCCTGGCTAATTTCTTATTTATTTAGTAGAGACGAGGTTTGCCATGTTG   60
          ||||| ||||| || ||||||||||| |||| ||||||||||||||| ||||||||||||
Db   1313 CCGGCCCCATACCCAGCTAATTTTTTATATTTTAGTAGAGATGGGGTTTGCCATGTTG  1372

Qy     61 ACCAGGGCTGGGTCTCGGAACTCTTGACCTCAGGTGATCGGCCTGCCTCAGCCTCCAAGTG  120
          ||| |||||||||||| |||||||||||||||| || |||| || || ||||||||||||
Db   1373 GCCAAGCTAGTCTCGAACTCCCGACCTCAAGTGATCACCCTGCCTCGGCCTCCCAAAGTG 1432

Qy    121 CTGGGATTATAGGCGTGAGCGCCGCGCCCAGC------CAACATTTTTTAAATACTGAA  174
          ||||||| |||| ||||||||  || | |||        ||||||||| ||| |||||
Db   1433 CTGGGATTACAGGTGTGAGCCACTGCACCTGGC CAACATTTTTTAAATACTGAA 1492

Qy    175 AAGTAGAGGGAAATAGTTATATAGTGTTACCCCATTACCCATCACTCAGTTCAACAGCTGGT  234
          ||||||||||| |||| |||   || |||||||| |||||||||||||||  |||||||
Db   1493 AAACAAAAAAAAACTTGACTGTGTCACTCGTGTTGTGTTCTCCTTACCTTGTATACTTCCA 1552

Qy    235 GACATATTTATTTCTTCTTATACCAGTACCGTACCTCCTGTTGTCTCTTTGTATACTTCCA  294
          || ||| |||||||| || ||| || ||| || ||| |||| ||||||| ||||
Db   1553 CAACTTCCCAGTGTTCTTGGATAAAGACCAAAATCCTTAACTTGCCAGGCGCGGTGGCT 1612

Qy    295 AAACCCAGATGACATTTTATCCCTAAATACTTTAGATAAAGGTGTTCTTTGAAAAAATC  354
          ||| ||| ||| ||||  |  | |||||||| | ||| || ||| |||| | |||||
Db   1613 CACACCTATCATCT------CAGCACTTTGGGAGGCCGAGGCGAGGCAGATCATGAAGTC 1665

Qy    355 ATAACCTCAGGACCACCAGCCTGCCAACATGGTGAAACCTGTCTGTGTTACTAAAAATACAAAA  414
          || |||||||||| || ||  ||||||||  |||||||| ||  | |||| |||||| ||||
Db   1666 AAGAGATTGAGACCATCCTGGCCAACATGGTGAAACCCCATCTCTACTAAAAATACAAAA 1725

Qy    415 ATTAGCTTGGCATGGTTCGTTGGGCACCTGTAATCCCAGCTACTCAGGAGGCTGAGGCAGGA  474
          ||||||| ||||||| |||| ||||||||||||||||||||| |||||||| |||||||||
Db   1726 ATTAGCTCGGTGTGGTGGCGTGCCTGTAGTCCCAGCTACTCGGGAGGCTGAGGCAGGA 1785

Qy    475 GAATCACTTGAATCCGGGAAGCAGAGATTGCAGTGAGCTGAGATCGAGCCTGGG  534
          |||||||||||| | |||||||||||||| ||||||||| |||| |||||||
Db   1786 GAATCACTTGAACCTGGGAGGCGGAGGTTGCAGTGAGCCGAGATCACGCCACTGCACTCC 1845

Qy    535 CGACAGAGACAGAAATGAAAACTCTGTCTCAAAAACAAACAAACAAAAA  584
          ||| ||||| ||||| |||||||||||||| |||||||||||||||
Db   1846 AGCCTGGTGACAGAGTAAGACTCCATCTCAAAAAAAAAAAAAAAAA 1895
```

OM nucleic - nucleic search, using sw model

GenCore version 5.1.4_p5_4578
Copyright (c) 1993 - 2003 Compugen Ltd.

Run on: March 29, 2003, 17:41:11 ; Search time 373.964 Seconds
                                          (without alignments)
                                          7242.188 Million cell updates/sec

Title: US-09-691-219-3_COPY_1_1000
Perfect score: 1000
Sequence: 1 cctgccacatgcctggcta.........caagagggctgcacctctga 1000

Scoring table: IDENTITY_NUC
               Gapop 10.0 , Gapext 1.0
Searched: 4883863 seqs, 135157058 residues

Total number of hits satisfying chosen parameters: 9767726

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
                 Maximum Match 100%
                 Listing first 45 summaries

Database :   Pending_Patents_NA_New:*
               1: /cgn2_6/ptodata/1/pna/PCT_NEW_COMB.seq:*
               2: /cgn2_6/ptodata/1/pna/US06_NEW_COMB.seq:*
               3: /cgn2_6/ptodata/1/pna/US07_NEW_COMB.seq:*
               4: /cgn2_6/ptodata/1/pna/US08_NEW_COMB.seq:*
               5: /cgn2_6/ptodata/1/pna/US09_NEW_COMB.seq:*
               6: /cgn2_6/ptodata/1/pna/US09_NEW_COMB.seq2:*
               7: /cgn2_6/ptodata/1/pna/US10_NEW_COMB.seq:*
               8: /cgn2_6/ptodata/1/pna/US10_NEW_COMB.seq2:*
               9: /cgn2_6/ptodata/1/pna/US60_NEW_COMB.seq:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	1000	100.0	11101	6	US-09-691-219-3	Sequence 3, Appli
2	229.8	23.0	177528	9	US-60-449-629-846	Sequence 846, App
3	226.6	22.7	630470	5	US-09-946-133-181	Sequence 181, App
4	217.6	21.8	128978	5	US-60-446-133-345	Sequence 345, App
5	211.8	21.2	126974	8	US-10-304-019-12	Sequence 12, Appl
6	210.2	21.0	170245	9	US-60-427-982-322	Sequence 322, App
7	205	20.5	1599662	5	US-09-947-911-108	Sequence 108, App
8	203.6	20.4	1554623	5	US-09-947-911-225	Sequence 225, Appl
9	202.6	20.3	1974088	5	US-09-947-911-73	Sequence 73, Appl
10	199.8	20.0	94720	1	PCT-US02-41414-1172	Sequence 1172, Ap
11	199	19.9	888463	5	US-09-947-911-117	Sequence 117, Appl
12	198.8	19.9	1059916	5	US-09-947-911-86	Sequence 86, Appl
13	198	19.8	1108317	8	US-10-017-161-2143	Sequence 2143, Ap
14	198	19.8	170245	9	US-60-427-982-322	Sequence 322, App
15	197.6	19.8	3037227	5	US-09-947-911-286	Sequence 286, App
16	197	19.7	4148350	5	US-09-947-911-107	Sequence 107, App
17	196.8	19.7	177528	9	US-60-449-629-846	Sequence 846, App
18	196.6	19.7	74545	6	US-09-949-002-606	Sequence 606, App
19	196.6	19.7	131860	6	US-09-949-002-730	Sequence 730, App
20	196.2	19.6	34827	8	US-10-272-810-10	Sequence 10, Appl
21	196.2	19.6	34827	8	US-10-272-810-10	Sequence 10, Appl
22	196.2	19.6	2593930	5	US-09-947-911-54	Sequence 54, Appl
23	195.4	19.5	86080	1	PCT-US02-40718-58	Sequence 58, Appl
24	195.2	19.5	4148350	5	US-09-947-911-107	Sequence 107, App
c 25	193.6	19.4	879306	5	US-09-947-911-194	Sequence 194, App
c 26	192.6	19.3	135284	9	US-60-446-133-149	Sequence 149, App
c 27	192.4	19.2	305349	5	US-09-947-911-44	Sequence 44, Appl
c 28	190.2	19.0	22789	5	US-09-947-911-8	Sequence 8, Appl
c 29	190	19.0	636160	5	US-09-947-911-21	Sequence 21, Appl
c 30	189.8	19.0	35425	8	US-10-017-161-2429	Sequence 2429, Ap
c 31	189.8	19.0	304905	1	PCT-US02-32700-1	Sequence 1, Appli
c 32	189.8	19.0	304905	1	PCT-US02-32700-1	Sequence 1, Appli
c 33	189.6	19.0	126974	8	US-10-304-019-12	Sequence 12, Appl
c 34	189.6	19.0	455237	1	PCT-US02-25201A-1	Sequence 1, Appli
c 35	189.4	18.9	2865598	5	US-09-947-911-26	Sequence 26, Appl
c 36	188.6	18.9	266910	9	US-60-449-155-1000	Sequence 1000, App
c 37	188.6	18.9	399733	5	US-09-947-911-290	Sequence 290, App
c 38	188.4	18.8	16815	5	US-10-017-161-1579	Sequence 1579, App
c 39	188.4	18.8	211975	9	US-60-449-155-992	Sequence 992, App
c 40	187.8	18.8	60194	6	US-09-949-002-656	Sequence 656, App
c 41	187.8	18.8	60195	6	US-09-949-002-697	Sequence 697, App
c 42	187.8	18.8	712680	5	US-09-947-911-209	Sequence 209, App
c 43	187.6	18.8	1108317	8	US-10-017-161-2143	Sequence 2143, Ap
c 44	187.4	18.7	888463	5	US-09-947-911-117	Sequence 117, Appl
c 45	187	18.7	119596	8	US-10-270-336-3	Sequence 3, Appli

ALIGNMENTS

RESULT 1
US-09-691-219-3
Sequence 3, Application US/09691219
; GENERAL INFORMATION:
; APPLICANT: WEI, Ming-Hui et al.
; TITLE OF INVENTION: ISOLATED HUMAN TRANSPORTER PROTEINS,
; TITLE OF INVENTION: NUCLEIC ACID MOLECULES ENCODING HUMAN TRANSPORTER PROTEINS,
; TITLE OF INVENTION: AND USES THEREOF
; FILE REFERENCE: CL000894
; CURRENT APPLICATION NUMBER: US/09/691,219
; CURRENT FILING DATE: 2000-10-19
; NUMBER OF SEQ ID NOS: 4
; SOFTWARE: FastSEQ for Windows Version 4.0
US-09-691-219-3

SEQ ID NO 3
LENGTH: 11101
TYPE: DNA
ORGANISM: Homo sapiens
FEATURE:
NAME/KEY: misc_feature
LOCATION: (1)...(11101)
OTHER INFORMATION: n = A,T,C or G

Query Match   100.0%; Score 1000; DB 6; Length 11101;
Best Local Similarity   100.0%;
Matches 1000; Conservative   0; Mismatches   0; Indels   0; Gaps   0;

Qy   1 CCTGCCACCATGCCTGGCTAATTTTTGTATTTTTAGTAGAGACGAGGTTTTGCCATGTTG 60
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1 CCTGCCACCATGCCTGGCTAATTTTTGTATTTTTAGTAGAGACGAGGTTTTGCCATGTTG 60

Qy   61 ACCAGGCTGGTCTCGAAACTCTTGACCTCAGGTGATCCGCCCGCCTCAGCCTCCCAAAGTG 120
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   61 ACCAGGCTGGTCTCGAAACTCTTGACCTCAGGTGATCCGCCCGCCTCAGCCTCCCAAAGTG 120

Qy   121 CTGGGATTATAGGCGTGAGCCGCCGCGCCCACCCAGCCAACATTTTTTAAAATACTGAAAGTAG 180
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   121 CTGGGATTATAGGCGTGAGCCGCCGCGCCCACCCAGCCAACATTTTTTAAAATACTGAAAGTAG 180

Qy   181 AGGGAATAGTTATATAGTGTACCCCATTTACCCCATCACTCAGTTTCAACAGCTGGTGACATA 240
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   181 AGGGAATAGTTATATAGTGTACCCCATTTACCCCATCACTCAGTTTCAACAGCTGGTGACATA 240

Qy   241 TTTTATTTCTTCTATACCAGTACCGTACCCCCACTGGGGATTATTTTTAAGGCAAAACC 300
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   241 TTTTATTTCTTCTATACCAGTACCGTACTCTCCCCACTGGGGATTATTTTTAAGGCAAAACC 300

Left column

```
Qy  301  AGATGACATTTTATCCTAAATACTTTAGATAAAAAGGTGTTCTTTGAAAAAAATCATAAACC  360
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  301  AGATGACATTTTATCCTAAATACTTTAGATAAAAAGGTGTTCTTTGAAAAAAATCATAAACC  360

Qy  361  TCAGGACCAGCCTGGCCAACATGGTGAAACCCTGTCTCTACTAAAAATACAAAAATTAGC    420
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  361  TCAGGACCAGCCTGGCCAACATGGTGAAACCCTGTCTCTACTAAAAATACAAAAATTAGC    420

Qy  421  TTGGCATGGTGGCGTGCGCCTGTAATCCCAGCTACTCAGGAGGCTGAGGCAGGAGAATCA   480
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  421  TTGGCATGGTGGCGTGCGCCTGTAATCCCAGCTACTCAGGAGGCTGAGGCAGGAGAATCA   480

Qy  481  CTTGAATCCGGGAAGCAGAGGTTGCAGTGAGCTGAGATCGCAGCCTGGGCGACAG       540
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  481  CTTGAATCCGGGAAGCAGAGGATTGCAGTGAGCTGAGATCGCAGCCTGGGCGACAG      540

Qy  541  AGACAGAAATGAAAACTCTGTCTCAAAAAAACAAACAAAAAAACACTATACATAAAAA     600
Db  541  AGACAGAAATGAAACTCTGTCTCAAAAAAACAAACAAAAAAACACTATACATAAAAA      600

Qy  601  TGAACAAATGATGCCACAATAGCCACCAGAGAATTTTATAAAATACAAGAATTCCCAGGCCCTGC  660
Db  601  TGAACAAATGATGCCACAATAGCCACCAGAGAATTTTATAAAATACAAGAATTCCCAGGCCCTGC  660

Qy  661  CCCAGACCTACTGAATCCTGGAAAATATTCAGGCTCCCACCACCCAGAGAATTCTGGTTCGGGTT  720
Db  661  CCCAGACCTACTGAATCCTGGAAAATATTCAGGCTCCCACCACCCAGAGAATTCTGGTTCGGGTT  720

Qy  721  GGTCTGATGGACGACCTGTAACCTGTTGTAACCTTCTCCAGGTAATGCTGAGCCT         780
Db  721  GGTCTGATGGACGACCTGTAACCTGTTGTAACCTTCTCCAGGTAATGCTGAGCCT        780

Qy  781  GCTGGTGCTCAGAGTAGACAGAGAAAAACCAGGGTGTCTGAGGGTTTTCCAGAAGA        840
Db  781  GCTGGTGCTCAGAGTAGACAGAGAAAAACCAGGGTGTCTGAGGGTTTTCCAGAAGA        840

Qy  841  AAACCAGAGTCCAGAGAAGGAGGAGAGGAGCACTCAGTGAGGTGAGGGGTGCAC         900
Db  841  AAACCAGAGTCCAGAGAAGGAGGAGAGGAGCACTCAGTGAGGTGAGGGGTGCAC         900

Qy  901  CTCACAATCCTCACTCCTGGCACCCCGTCTCCTACAAGAGATGAGGACCCAGAGCTGAAAGAGCCCCTT  960
Db  901  CTCACAATCCTCACTCCTGGCACCCCGTCTCCTACAAGAGATGAGGACCCAGAGCTGAAAGAGCCCCTT  960

Qy  961  CCTGTCCCCAGTGGTGTTGGGGCAAGAGGGCCTGCACCTCTGA  1000
Db  961  CCTGTCCCCAGTGGTGTTGGGGCAAGAGGGCCTGCACCTCTGA  1000
```

RESULT 2
US-60-629-846
Sequence 846, Application US/60449629
; GENERAL INFORMATION:
; APPLICANT: CARGILL, Michele
; TITLE OF INVENTION: GENETIC POLYMORPHISMS ASSOCIATED WITH
; TITLE OF INVENTION: CANCER, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001449
; CURRENT APPLICATION NUMBER: US/60/449,629
; CURRENT FILING DATE: 2003-02-04
; NUMBER OF SEQ ID NOS: 2712
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 846
; LENGTH: 177528
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)..(177528)
; OTHER INFORMATION: n = A,T,C or G, or insertion/deletion polymorphism (see Tables 1
US-60-449-629-846

Query Match   23.0%;   Score 229.8; DB 9; Length 177528;
Best Local Similarity 61.3%; Pred. No. 9.7e-26;
Matches 403; Conservative 0; Mismatches 252; Indels 2; Gaps 2;

Right column

RESULT 3
US-09-947-911-181/c
Sequence 181, Application US/09947911
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig
; TITLE OF INVENTION: COLLECTION OF SINGLE NUCLEOTIDE POLYMORPHISM (SNPs) LOCATED ON
; TITLE OF INVENTION: X, METHODS OF DETECTION, AND USES THEREOF
; FILE REFERENCE: CL001300
; CURRENT APPLICATION NUMBER: US/09/947,911
; CURRENT FILING DATE: 2001-09-07
; NUMBER OF SEQ ID NOS: 368
; SEQ ID NO 181
; LENGTH: 630470
; TYPE: DNA
; ORGANISM: HUMAN
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)..(630470)
; OTHER INFORMATION: n = A,T,C or G
US-09-947-911-181

Query Match   22.7%;   Score 226.6; DB 5; Length 630470;
Best Local Similarity 64.9%; Pred. No. 2e-25;
Matches 421; Conservative 0; Mismatches 209; Indels 19; Gaps 60;

```
Qy    1  CCTGTCCCCAGTGGTGTTGGGGCAAGAGGGCCTGCACCTCTGA ...
```

```
RESULT 4
US-60-446-133-345/c
Sequence 345, Application US/60446133
GENERAL INFORMATION:
  APPLICANT: Wyeth
  APPLICANT: Burczynski, Michael E
  APPLICANT: Twine, Natalie E
  APPLICANT: Dorner, Andrew E
  APPLICANT: Trepicchio, William E
  TITLE OF INVENTION: Method for Monitoring Anti-Cancer Drug Activities
  FILE REFERENCE: AM101055-2
  CURRENT APPLICATION NUMBER: US/60/446,133
  CURRENT FILING DATE: 2003-02-11
  NUMBER OF SEQ ID NOS: 386
  SOFTWARE: Patentin version 3.1
  SEQ ID NO 345
  LENGTH: 128978
  TYPE: DNA
  ORGANISM: Homo sapiens

  Query Match       21.8%;  Score 217.6;  DB 9;  Length 128978;
  Best Local Similarity   63.2%;  Pred. No. 6e-24;
  Matches  373;  Conservative    0;  Mismatches  204;  Indels  13;  Gaps  2;

Qy      1  CCTGCCACCATGCCTGGCTAATTTCTTATATTTTTAGTAGAGACGGGGTTTCACCATGTTG     60
              |||||||||  ||  |||||||| ||||  |||||||||||||||||| ||||||  ||||||||
Db  72729  CCGGCCCCATACCCAGCTAATTTTTTATATTTTTAGTAGAGATGGGGTTTTGCCATGTTG  72670

Db  99673  CCTGCCACCACCTGGCTGCTCGAACTCTTGACCTCAGGTGATCC GCCTGCCTCAGCCTCCCAAAGTG  99614
Qy     61  ACCAGGCTGGTCTCGAACTCTTGACCTCAGGTGATTC                                  119

Db  99613  GCCAGCTGGTCTGGAACTCCTGACCTCAGGTGATCACCCACCTCAGCCTCCCAAAGT  99554
Qy    120  GCTGGGATTATAGGCGTGAGCCCGCGACCCAGCCAACATTTTTAAAATACTGAAAAGTA    179

Db  99553  GCTGGGTTTACAGGCATGAGCCACCACGCCCGGCCCCGTGTTT  TCCATCTTT  99504
Qy    180  GAGGGAATAGTTATAGTGTACCCCATTTTACCCATCACACTGCAGTTCAACAGCTGTGACAT    239

Db  99503  CAAGGTACTGGAATAATTCAAATGCCATTTTTCCCAGCCTTACAGTTCTAGTCGTGGCATAA  99444
Qy    240  ATTTATTTTCTTCTATACCAGTACCGTATCTCCCCACTGGGGATTTATTTTAAGGCAAAACC    299

Db  99443  AAACCT------TCTCTAAGCACCCTGCATAGAACCACATTTTAGAGGCGCAGGCGTGGTG  99390
Qy    300  CAGATGACATTTTTATCCCTAAATACTTTAGATAAAAGGTGTTCTTGAAAAAATCATAAC    359

Db  99389  GCTCACGCCTGTAATCCCAGCACTTGGGAGGCCAAGGCGGGATCACGAGGGTCAGGAG  99330
Qy    360  CTCAGGACCAGCCTGGCCAACANGTGAAAACCCTGTCTCTGTACTAAAAAATACAAAAAATTAG    419

Db  99329  TTTGAGACCAGCCTGGCCAACATGGTGAAACCCCATCTCTACTAAAAATACAAAAAATTAG  99270
Qy    420  CTTGGCCATGGTCGTGTGGGCACCTGTAATCCCAGCTACTCAGGAAGGCTGAGGCAGGAGAATC    479

Db  99269  CTGGGCGTGGTGGCGGGCACCTGTAATCCCAGCTACTCGGGAGGCTGAGGCAGGAATC  99210
Qy    480  ACTTGAATCCGGGAAGCAGAGATTGCAGTGAGCTGAGATCGCAGCCTGGGCGACA    539

Db  99209  GCTTGAACCTGGGAGGCGGAGGTTGCAGTGAGCCGAGATTGAGCCACTGCACTCCAGTCT  99150
Qy    540  GAGACAGAAATGAA-ACTCTGTCTCAAAAAACAAACAAACAAAAAAC -CACTATACATAA    597

Db  99149  GGGTGACAGAGCAAGAACTCCGTCTCAAAAAAAAAAAAAAAAAAGAATCACATTTTTTAGCTAG  99090
Qy    598  AAATGACAATGAATGCCACCAGTAGCACCAGAGAATTTTATAAATACACAGA    646

Db  99089  AGAGAAAAAATGGTGGCAGAAGAACACAGGATAGCCATTAATTAAAGTCTCAGA  99041
```

```
RESULT 5
US-10-304-019-12/c
Sequence 12, Application US/10304019
GENERAL INFORMATION:
  APPLICANT: Nicholas M. Dean
  APPLICANT: C. Frank Bennett
  APPLICANT: Kenneth W. Dobie
  TITLE OF INVENTION: MODULATION OF HEPATOCYTE GROWTH FACTOR RECEPTOR EXPRESSION
  FILE REFERENCE: PTS-0043
  CURRENT APPLICATION NUMBER: US/10/304,019
  CURRENT FILING DATE: 2002-11-23
  NUMBER OF SEQ ID NOS: 147
  SEQ ID NO 12
  LENGTH: 126974
  TYPE: DNA
  ORGANISM: H. sapiens
  FEATURE:
  US-10-304-019-12

  Query Match       21.8%;  Score 211.8;  DB 8;  Length 126974;
  Best Local Similarity   63.8%;  Pred. No. 4.1e-23;
  Matches  391;  Conservative    0;  Mismatches  207;  Indels  15;  Gaps  4;

Qy      1  CCTGCCACCATGCCTGGCTAATTTCTTATATTTTTAGTAGAGACGGGGTTTCACCATGTTG     60
Db 122866  CCTGCCACCATGCCCGGCTAATTTTTTGTATTTTTAGTAGAGATGGGGTTTCTCCATGTTG 122807

Qy     61  ACCAGGCTGGTCTCGAACTCTTGACCTCAGGTGATCCGCCTCCCAGCCTCCCAAAGTG    120
Db 122806  GCCAGGCTGGTCTCTCGAACTCCTGACCTCAGGTGATCCACCTGCCTCGGCCTCCCAAAGTG 122747

Qy    121  CTGGGATTATAGGCGTGAGCCGCCGCGCCCAGC-------CAACATTTTTTAAAATACTGAA    180
Db 122746  CTGGGATTACAGGCATGAGCCACCGCGCCCGGCCCCAGCCAACATTTTTTATTTAAACCATAA 122687

Qy    181  AGGGAATAGTTATAGTGTACCCCA------TTTACCCATCACTCACTGTTTCTTATTAAACAGCTG    232

Qy     61  ACCAGGCTGGTCTCGAACTCTTGACCTCAGGTGATGATCCGCCTGCCTCAGCCTCCCAAAGTG    120
Db  72669  GCCAAGCTAGTGTCGAACTCCTGACCTCAAGTGATCACCCTGCCTCAGCCTCCCAAAGTG  72610

Qy    121  CTGGGATTATAGGCGTGAGCCGCCGCGCCCAGC-------CAACATTTTTTAAAATACTGA    174
Db  72609  CTGGGATTACAGGTGTGAGCCACTGCGCCCGGCCACCCAGCCAACATTTTTTAAAATACTGA  72550

Qy    175  AAGTAGAGGGAATAGTTATAGTGTACCCCATTTTACCCATCACACTGCAGTTCAACAGCTGGT    234
Db  72549  AAACAAAAAAAACTTGACTGTGTCACTCGTGTCTCCTCACCTTGTATACTTCCA  72490

Qy    235  GACATATTTATTTCTTCTATACCAGTACCGTATCTCCCCACTGGGGATTTATTTTAAGGCA    294
Db  72489  CAACTTCCCCAGTGTTCTTTGGATAAAAGACCAAAATCCTTAACTGGGGGATTATTTTAAGGAGTC  72430

Qy    295  AAACCCAGATGACATTTTTATCCCTAAATACTTTAGATAAAAGGTGTTCTTGAAAAAAATC    354
Db  72429  CACACCTATCATCT-------CAGCACTTGGGAGGCCGAGGCAGCAGGATCATGAGAGTC  72377

Qy    355  ATAACCTCAGGACCACCCTGCCACACATGGTGAAACCCTGTCTGTACTAAAAAATACAAAA    414
Db  72376  AAGAGATTGAGACCATCCTGGCCAACATGGTGAAACCCCATCTCTACTAAAAATACAAAA  72317

Qy    415  ATTAGCTTGGCATGGTGCGTGGGCACCTGTAATCCCAGCTACTCAGGAAGGCTGAGGCAGGA    474
Db  72316  ATTAGTCGTGGTGGCGTGCGCCTGTAGTCCCAGCTACTCGGGAGGCTGAGGCAGGA  72257

Qy    475  GAATCACTTGAATCCGGGAAGCAGAGATTGCAGTGAGCTGAGATTGCAGTGAGCCGAGGCTGGG    534
Db  72256  GAATCACTTGAACCTGGGAGGCGGAGGTTGCAGTGAGCCGAGATCACGCCACTGCACTCC  72197

Qy    535  CGACAGAGACAGAAATGAAACTCTGTCTCAAAAAACAAAACAAAACAAAAAAA    584
Db  72196  AGCCTGGTGACAGAGTAAGACTCCATCTCAAAAAAAAAAAAAAAAAAA  72147
```

```
Db 122686 AGTATAATATAGTTCACGTTGCTTGACATGTTGGAGTAGGGAGGAGGGATTAGGGACCCATAA 122627
                                                           ||
Qy    233 GTGACATATTTATTTCTTCTATACCAGTAGTTGGAGTAGGGAGGAGGGAT-TATTTTAAG 291
                                                           ||
Db 122626 GATTGAAATGGGGACTGTGCTGTGGGAGGTTTTTAAATGTTAGAAAGTTAGACAGGACG 122567
                                                           ||
Qy    292 GCAAAACCCAGATGACATTTATTCCTTAAAAATACTTTAGATAAAAGGTGTTCTTTGAAAA 351
                                                           ||
Db 122566 GGACAAGGTGGTTCACATTCTGTAATCCAGCACTTTGGGGAGGCCGGGGGCAGGATCAC 122507
                                                           ||
Qy    352 A-----TCATAAACCTCAGGACGACCTGGCCCAACATGGTGAAACCCTGTCTCTACTAAAA 406
                                                           ||
Db 122506 TTGAGGTCAGAGTTCGGGACCAGCCTGGCCAACATGGTGAAACCCTGTCTCTACTAAAA 122447
                                                           ||
Qy    407 ATACAAAAATTAGCTGGGCATGGTGGCGTGCGCCTGTAATCCCAGCTACTCAGGAAGCTG 466
                                                           ||
Db 122446 ATACAAAAAATTAGCCAGGCATGGTGGTGTGCGCACCTGTAATCCCAGCTACTCGGGAGGCTG 122387
                                                           ||
Qy    467 AGGCAGGAGAATCACTTGAATCCGGGAAGCAGAGATTGCAGTGAGCTGAGATTGCAGTCG 526
                                                           ||
Db 122386 AAGCAGGAGAACATCGCTTGAACCTGGGAGGTGGCAGTGAGCTGAGATCGCCGCCCC 122327
                                                           ||
Qy    527 AGCCTGGGCGGGACAGAGAGCAACAGA-AATGAAACTCTGTCTCAAAAAACAAACAAAAAAAC 585
                                                           ||
Db 122326 TGCACTCCAGCTTGGGCGATAGAGTGAGACTCCGTCTCAAAAAAAAAAAAAAAAGAAAAGTG 122267
                                                           ||
Qy    586 CACTATACATAAA 598
                                                           ||
Db 122266 CGTTAGACTGAAA 122254
```

RESULT 6
US-60-427-982-322/c
Sequence 322, Application US/60427982

Query Match   21.0%; Score 210.2; DB 9; Length 170245;
Best Local Similarity 61.4%; Pred. No. 6.5e-23;
Matches 374; Conservative 0; Mismatches 228; Indels 7; Gaps 2;

```
Qy      2 CTGGCCACCATGCCCGCTCTGGCTAATTTCTTATTTTTAGTAGAGACGAGGGTTTGCCATGGTTGA 61
            ||||||| || ||||||||||||| |||| |||| |||||||||||| ||||||||||||
Db  64533 CCGCCACCACGCCCGCTAATTTTGTATTTTTAGTAGAGATGGGGTTTCACCATGTTGG 64474

Qy     62 CCAGGCTGGTCTCGAACTCTTGACCTGTGATCCGCCCGCCTCCCAGCCTCCCAAAGTGC 121
           |||||||||||| |||||| ||||||  |||| |||||||| ||||||||||||||
Db  64473 TCAGGCTGGTCTTGAACTCCTGACCTCAGGTTATCGCCCGCCTCAGCCTCCCAAAGTGT 64414

Qy    122 TGGGATTATAGGCGTGAGCGCCGCGCCCAGCCAA----CATTTTTAAATACTGAAAG 177
           | ||||||| |||||||||||||||||||||||    |||||||||||| |||||
Db  64413 TAGGGTTACAGGCGTGAGCCATTGCGCCCACCAGCCAATTTGCTTTTTTTTAAGAAAAAAAAT 64354

Qy    178 TAGAGGGAAATAGTTATATAGTGTACCCCATTTACCCATCACTCAGTTTCAACAGCTGGTGAC 237
           || || ||| ||  ||| ||  |||||| || |||| |||||| |||||||||| ||||
Db  64353 TAAGAAAATGCCTGGAACTTCTAAGAAAAAAAAACCTGAATGAGTTAAAAAATTAGAC 64294
```

RESULT 7
US-09-947-911-108/c
Sequence 108, Application US/09947911

Query Match   20.5%; Score 205; DB 5; Length 1599662;
Best Local Similarity 62.8%; Pred. No. 2.1e-22;
Matches 371; Conservative 0; Mismatches 210; Indels 10; Gaps 3;

```
Qy      4 GCCACCATGCCCGCTAATTTCTTATTTTTAGTAGAGACGAGGGTTTGCCATGGTTGACC 63
          ||||||||||||   |||||| ||||| ||||||||||| || ||||| |||||||| ||
Db1386342 GCCACCATGCCCAGCGTAATTTAATTTTTGTGTTTTAGTAGAGACACAGGGTTTCACCATGTTGGCC 1386283

Qy     64 AGGCTGGTCTCGAACTCTTGACCTGTGATCCGCCCTGCCTCCCAGCCTCCCAAAGTGCTG 123
          ||| ||||||| |||||  ||  || |||  |||| |||| |||| ||||||||||||||
Db1386282 AGGGTGGTCTCAGACTCCTCTGATGTCAGGTGATCTACCGCGCCTCGGCCTCCCAAAGTGCTA 1386223

Qy    124 GGATTATAGGCGTGAGCGCCGCGCCCACCCAGCCAACACATTTTTAAATACTGAAAGTAGAGG 183
          ||||| ||| ||||||| || |||||| || ||| |  |||| |||||| || |||
Db1386222 GGATTACAGGTGTGAGCCACCGCGCCCGGCCTACAAATTTGTTTTTATAATTATAGGTA 1386163

Qy    184 GAAT-AGTTATAGTGTACCCCATTTACCCATCACTCAGTTTCAACAGCTGGTGACATATT 242
          | ||  ||  || | | | |  ||  |||| | |    |||    |  | ||||
Db1386162 TATTCAGATAACTCTTTTCTTCTGGAGCAAGCTTTGGTGGTTTGTGGGCCTGGAAGGAATT 1386103

Qy    243 TATTTCTTCTATACCAGTACCGTACTCTCCCCACTGGGATTATTTTTAAGGCAAAACCCAG 302
          | | || | ||| | |  ||  ||||  ||  |  |||| |||||||| ||||  || ||
Db1386102 TGTCTCATCTAAACTACCAAGTTTTTGATATAAAGTTTTTGGGCCAGGTGGGGTGGGCT 1386043
```

```
Qy        303  ATGACATTTTATCCCTAAATACTT-TAGATAAAGGTGTTCTTTGAAAAAAATCATAACCT  361
Db    1386042  CACGCCTGTAAACCCACCACTTTGGGAGGCCTTAGGCGAGACGGATCTCGAGGTCAGGAGTT  1385983

Qy        362  CAGGACCAAGCCTGGCCCAACATGGTGAAAACCCTGTCTGTACTAAAAATACAAAAATTAGCT  421
Db    1385982  CAAGACCAAGCCTCAGGCCCAACATAGTGAAAACCCTGTCTCTACTAAAAATACAAAAATTAGCC  1385923

Qy        422  TGGCATGGTGGTGTGGGCACCTGTAATCCCAGCTACTCAGGAAGGTGACGAGGAAATCAC  481
Db    1385922  GGGCATGGTGGACAGTGCCTGTAGTACCGGCTACTCGGGAGGCTGAGGCAGGAGAATCGC  1385863

Qy        482  TTGAATCCGGGAAGCAGAGATTGCAGTGAGCTGAGAT----------TGCAGTCGAGCGCTGG  533
Db    1385862  TTGAAACCGGGAGGCGGAGGTTGCAGTGAGCCGAGATCGTGCCACTCCACTCTAGCTGG  1385803

Qy        534  GCGACAGAGAGACAGAAATGAAACTCTGTCTCAAAAAAATAAAGAAAATAAAACAAACAAAAAA  584
Db    1385802  GCAACAGAGTGAGACTTCATCTCAAAAAAATAAAGAAAATAAAAATAA  1385752
```

RESULT 8
US-09-947-911-225/c
; Sequence 225, Application US/09947911
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig
; TITLE OF INVENTION: COLLECTION OF SINGLE NUCLEOTIDE POLYMORPHISM (SNPs) LOCATED ON CH
; TITLE OF INVENTION: X, METHODS OF DETECTION, AND USES THEREOF
; FILE REFERENCE: CL001300
; CURRENT APPLICATION NUMBER: US/09/947,911
; CURRENT FILING DATE: 2001-09-07
; NUMBER OF SEQ ID NOS: 368
; SEQ ID NO 225
; LENGTH: 1554623
; TYPE: DNA
; ORGANISM: HUMAN
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)..(1554623)
; OTHER INFORMATION: n = A,T,C or G
US-09-947-911-225

Query Match   20.4%; Score 203.6; DB 5;   Length 1554623;
Best Local Similarity   61.6%; Pred. No. 3.3e-22;
Matches 377; Conservative   0; Mismatches 229; Indels 6; Gaps 3;

```
Qy          1  CCTGCCACCATGCCTGGCTAATTTTCTTTATTTTAGTAGAGACGAGGTTTTGCCATGTTG  60
Db    1429531  CATGCCACCACGCCCACCTAATTTGTGTATTTTTAGTAGAGACAGGGTTTCATCATGTTG  1429472

Qy         61  ACCAGGCTGGTCTCGAACTCTTGACCTCAGGTGGATCCGCCGCCTCAGCCTCCCAAAGTG  120
Db    1429471  GCCAGGCTGGTCTTGAACTCCTGATTTCAAGTGATCTCCCTGCCATTGGCCTCCTGAAGTG  1429412

Qy        121  CTGGGATTATAGGCGTGAGCCGCGCGCCACCCAGCCCAACACATTTTTAAATACTGAAAAGTAG  180
Db    1429411  CTTGGGATTACAGGCATGAGTCACTGCACCCGGCCTCCCATGTTGGTCTTTTATAGCCACA  1429352

Qy        181  AGGGAAATAGTTATAGTGTACCCCATTTACCCACTCAGTTGCAACACTGTTGACATA  240
Db    1429351  GTCACACTCTCTTGCCTACTACCCAGTA-TCATCTCTAATCCCTGAAAACAACTGATATA  1429293

Qy        241  TTTATTTCTTCTATACCAGTACGTACTCTCCCCACTGGGATTATTATTTTAAGGCAAAACCC  300
Db    1429292  TTCCCCATCTCTATATAAATTTTCTCATTTCAAGAATGCTAGAATGGGCTGGGCATGGTG  1429233

Qy        301  AGATGACATTTTATCCTCTAAAATACTTTAGATAAAGGTGTTCTTTGAAAA---AAATCAT  356
Db    1429232  GCTTGCACCTGTAATTCCAGCACTTCGGAGAAGCCTGAGGCAGGTGGATTACCTGAGGTCAG  1429173

Qy        357  AACCTCAGGACCAGCCCTGGCCAACATGGTGAAACCCTGTCTGTACTAAAAATACAAAAAT  416
```

```
Db    1429172  GAGTTCTAGAACAGCCTGCCCAACATGGTGCAAAACCTTGTCGTCTACTAAAAATACAAAAAT  1429113

Qy        417  TAGCTTGGCATGGTCCGTTGGGCGGCCACCTGTAATCCCAGCTACTCAGGAAGGCTGAGGCAGGAGA  476
Db    1429112  TAGCTGGGGCGGTGGGCGGTGGCACACCTGTAGTCCCAGCTACTAGGGAGGCCGAGGCAGGAGA  1429053

Qy        477  ATCACTTGAATCCGGGAAGCAGAGAGATTGCAGTGAGCTGAGATTGCAGTCGAGCCTGGGCG  536
Db    1429052  ATTGCTTGAATCTGGGAGGCAGAGGTTGCAGTTAGCCGAGATCGCACCACTGCACTCCAG  1428993

Qy        537  ACAGAGACAGA-AATGAAAACTCTGTCTCAAAAAAACAAACAAAACAAAAAAAACCACTATACAT  595
Db    1428992  CCTGGGCGACAGAGCGAGACTCCATCTCAAAAATAAAATAAAAATACATAAAAAAG  1428933

Qy        596  AAAAATGAAACAA  607
Db    1428932  AATGTTATATAA  1428921
```

RESULT 9
US-09-947-911-73
; Sequence 73, Application US/09947911
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig
; TITLE OF INVENTION: COLLECTION OF SINGLE NUCLEOTIDE POLYMORPHISM (SNPs) LOCATED ON
; TITLE OF INVENTION: X, METHODS OF DETECTION, AND USES THEREOF
; FILE REFERENCE: CL001300
; CURRENT APPLICATION NUMBER: US/09/947,911
; CURRENT FILING DATE: 2001-09-07
; NUMBER OF SEQ ID NOS: 368
; SEQ ID NO 73
; LENGTH: 1974088
; TYPE: DNA
; ORGANISM: HUMAN
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)..(1974088)
; OTHER INFORMATION: n = A,T,C or G
US-09-947-911-73

Query Match   20.3%; Score 202.6; DB 5;   Length 1974088;
Best Local Similarity   62.9%; Pred. No. 4.3e-22;
Matches 363; Conservative   0; Mismatches 209; Indels 5; Gaps 3;

```
Qy          1  CCTGCCACCATGCCTGGCTAATTTTCTTTATTTTAGTAGAGACGAGGTTTTGCCATGTTG  60
Db     368267  CCCGCCACCACACCGCCTAATTTTGTTATTTTTAGTAGAGATGGGGTTTCACCATGTTG  368326

Qy         61  ACCAGGCTGGTCTCGAACTCTTGACCTCAGGTGATCCGCCGCCTCAGCCTCCCAAAGTG  120
Db     368327  TCCAGGCTGGTCTTGAACTCCTGACCTCAGGTGATCCAGCCGCCTCAGCCTCCCAAAGTG  368386

Qy        121  CTGGGATTATAGGCGTGAGCCGCGCGCCAGCCCAACACATTTTTAAATACTGAAAGTAG  180
Db     368387  CTGGGATTACAGGTGTGAGCCACCACGCCCAGCACTTCATACCATGAGATATGTCT  368446

Qy        181  AGGGAAATAGTTATAGTGTACCCCATTTACCCACTCAGTTTCAACACAGCTGGTGTGACATA  240
Db     368447  TCTGGCCCTACACTTTCAGTTCAAAACTCTGGTAAGTCATCAC-ACTGGGGTAGAAATA  368505

Qy        241  TTTATTTCTTCTATACCAGTACCGTACTCTCCCCACTGGGATTATTATTTTAAGGCAAAACCC  300
Db     368506  TTCCTCGAAGCTATTCCTACACCAAGAAGT---CTATGAATAACAGCCCGCGCGTGTGGC  368562

Qy        301  AGATGACATTTTTATCCTCTAAATACTTTAGATAAAGGTGTTCTTTTGAAAAAAATCATAACC  360
Db     368563  TCACGCCAGTAATCCCAGCACTTGGCAGGCCGAAGCTGACAGATCATGAGGTCAGGAGT  368622

Qy        361  TCAGGACCAGCCTGGCCAACATGGTGAAAACCCTGTCTGTACTAAAAATACAAAAAATTAGC  420
Db     368623  TCAAGACCAACCTGACCAATATGACGAAAACCCCATCTCTACCAAAAATACAAAAAATTAGC  368682

Qy        421  TTGGCATGGTGGTGGGCACCTGTAATCCCAGCTACTCAGGAAGCTGAGGCAGGAGAATCA  480
```

```
Db 368683 CGGGTGTGGTGACAGGCGGCCGGGAAGCAGAGAGGACGGGCGCCTGTAGTCCCAGCTACTTGGGAGGCTGAGGCAGGAGAATAG 368742
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy    481 CTTGAATCCGGGAGGCGGAGGATTGCAGTGAGCTGAGATTGCGTGGCCGCGACAG 540
          |||||| |||||||||||||| |||||||||| ||| || |||| |||
Db 368743 CTTGAACCCGGGAGGCGGAGGGTTGCAGTGAGCCGGAGATTGTGCCACTCCAGCCTG 368802
Qy    541 AGACAGA-AATGAAACTCTGTCTCAAAAAACAAACAA 576
          |||||||  |||| |||||||||||||||||||||||
Db 368803 GGTGACAGAGTGAGACTCTGTCTCAAAAAAAAAAAAAA 368839
```

RESULT 10
PCT-US02-41414-1172/c
Sequence 1172, Application PC/US0241414

; Sequence 1172, Application PC/US0241414
; GENERAL INFORMATION:
; APPLICANT:
;   APPLICANT: Morris, David
;   APPLICANT: Engelhard, Eric
; TITLE OF INVENTION: NOVEL COMPOSITIONS AND METHODS FOR CANCER
; FILE REFERENCE: A-71249/RMS/DCF
; CURRENT APPLICATION NUMBER: PCT/US02/41414
; CURRENT FILING DATE: 2002-12-26
; PRIOR APPLICATION NUMBER: US 09/747,377
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: US 09/798,586
; PRIOR FILING DATE: 2001-03-02
; NUMBER OF SEQ ID NOS: 1613
; SOFTWARE: Patentin version 3.1
; SEQ ID NO 1172
; LENGTH: 94720
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (23326)..(24418)
;   OTHER INFORMATION: "n" at positionsm 23326 thru 24418 can be any base
; FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (87811)..(87830)
;   OTHER INFORMATION: "n" at positions 87811 thru 87830 can be any base
PCT-US02-41414-1172

Query Match   20.0%; Score 199.8; DB 1; Length 94720;
Best Local Similarity   60.6%; Pred. No. 2.4e-21;
Matches 391; Conservative 0; Mismatches 232; Indels 22; Gaps 3;

```
Qy      5 CCACCATGCCTGGCTAATTTTTCTTTATTTTTAGTAGRAGACGAGGTTTTTGCCATGTTGACCA 64
          |||||| |||| |||||||| |  ||  |||||||||||||| |||| ||||||||||||||
Db  56870 CCACCACGCCTGGCTAAATTTTTTGTATTTTTAATAGAGACGGGGTTTTTACCATGTTGCCA 56811
Qy     65 GGCTGGTCTCGAACTCTTGACCTCAGGTGATCCGCCCGCCTCCCAGCCTCCCAAAGTGCTGG 124
          |||||||||| |||||||||||||| |||||||||||||||||||||||||||||||||
Db  56810 GGCTGGTCTCAAACTCCTGATCTCAAGTGATCCACCCACCTCAGCCTCCCAAAGTGCTGG 56751
Qy    125 GATTATAGGCGTGAGCCGCCGCCACCAGCC----------AACATTTTTAAATAC 170
          |||| ||||||||||||  ||  |||| ||          |||| || || |||
Db  56750 GATTACAGGCGTGAGCCACCGTGCCCGGCCAGGAGCGCCACTCTTAAAGAGATCCAGTCTC 56691
Qy    171 TGAAAAGTAGAGGGAATAGTTATAGTGTACCCCCATTCACCATCACTCAGTTTCACAGC 230
          ||||| | |||||||||| ||| |||  |||||  |||| |||||||| |||||||
Db  56690 TGAGCAAAACTCACAGAAGTCAGGTAATGGTTGTCTTTACTTGAAAACCTTTAATTCTGTTA 56631
Qy    231 TGGTGACATATTTATTTCTCTATACCAGTACCGTACTCCCCACTGGGGATTATATTTAA 290
          |||| ||| || || |||||||||||||||||||||||||||||||||||||||||||
Db  56630 AAAAAAAAAAGACTTGCTTAGATGCTTAAACATAAAAGAGACTAA---AGATTTGGCTGG 56574
Qy    291 GGCAAAAACCAGATGACATTTTTATCCTAAATACTTTAGAAAGGTGTTCTTTGAAAA 350
          ||||||||||| |  |||| || ||||  ||| |||||||||||||||| |||||
Db  56573 GCCCGGTGGCTCAAGCCTTTAATCCCAGCACTTTGGGAGGCCGAAGTGACAGTCACACAA 56514
Qy    351 AATCATAAACCTTCAGGACCTGGCCAACATGGTGAAACCCTGTCTGTCTAAAAATAC 410
          |||| |||  |||||||||||||||||||||||||||||||||| || ||||||| |
Db  56513 GGTCAGGAGTTAGAGACCAGCCTGGCCAACATGATGAAACCCCGGTCTCTATTAAAAATAC 56454
```

```
Qy    411 AAAAATTAGCTTGGCATGGTGGTCGTTGGGCCACTGTAATCCCAGCTACTCAGGAAGCTGAGGC 470
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  56453 AAAAATTAGCTTGGCGTGGTGGTGGGCGCACCTGCAGTGGTCCCAGCTACTCGGGAAGCTGAGGC 56394
Qy    471 AGGAGAATCACTTGAATCCGGGAGGCAGAGAGAGTTGCAGTGAGCTGAGATGCCGGGAACTGCGAGC 530
          ||||||| || ||||||||||||||||||||||||||||||||||||||||||| |||||||||||
Db  56393 ACGAGAATTGCTTGAAACCGGGAGGCGGAGGTTGCAGTGAGCCGAGATCACGGCCACTCGA 56334
Qy    531 TGGGCGACAGAGCGAGACTCTGTCTCAAAAAACAAACAAAAACAAAAAAAACCACTA 590
          |||||| ||||| |||| ||||||||||||||||||||||||||||||||||||||||
Db  56333 GCCTGGTGACAG-----AGGGAGACTCCATCTCAAAAAAAAAAAGAATTGAAAATTCTT 56279
Qy    591 TACATAAAAATGAACAATGATGCCACAATAGCACCAGAGAATTT 635
          |||| ||||||||||| |||||||||| |||||||||| ||
Db  56278 CATTAAAAATCTTTATATTCTGAAGAAAACCAAAGCTAAAAAATTT 56234
```

RESULT 11
US-09-947-911-117/c
Sequence 117, Application US/09947911

; Sequence 117, Application US/09947911
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig
; TITLE OF INVENTION: COLLECTION OF SINGLE NUCLEOTIDE POLYMORPHISM (SNPs) LOCATED ON
; TITLE OF INVENTION: X, METHODS OF DETECTION, AND USES THEREOF
; FILE REFERENCE: CL001300
; CURRENT APPLICATION NUMBER: US/09/947,911
; CURRENT FILING DATE: 2001-09-07
; NUMBER OF SEQ ID NOS: 368
; SEQ ID NO 117
; LENGTH: 888463
; TYPE: DNA
; ORGANISM: HUMAN
; FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (1)..(888463)
;   OTHER INFORMATION: n = A,T,C or G
US-09-947-911-117

Query Match   19.9%; Score 199; DB 5; Length 888463;
Best Local Similarity   62.4%; Pred. No. 1.7e-21;
Matches 410; Conservative 0; Mismatches 215; Indels 32; Gaps 5;

```
Qy      5 CCACCATGCCTGGCTAATTTTTCTTTATTTTTAGTAGRAGACGAGGTTTTTGCCATGTTGACCA 64
          |||||| |||| |||||||| | ||  ||||||||||||| |||| ||||||||||||||| ||
Db 594911 CCACCACGCCTGACTACTTTTTGTATTTTTAGTAGAGACGGGGTTTCGCCATGTTGGCCA 594852
Qy     65 GGCTGGTCTCGAACTCTTGACCTCAGGTGATCCGCCCGCCTCAGCCTCCCAAAGTGCTGG 124
          |||||||||| |||||||||||||| ||||||| ||| |||||||||||||||||||||||
Db 594851 GGCTGGTCTTGAACTCCTGACCTCGTGATCCACCCGCCTCAGTTGGCCTCCCAAAGTGCTGG 594792
Qy    125 GATTATAGGCGTGAGCCGCCGCCGCCGCCACCAGCCAACATTTTTAAATTACTAGAGG 184
          |||| |||||||||| |||| || ||||| ||||||||||||||||||||||||||||
Db 594791 GATTACAGGCGTCGAGCCACCACGCCCGGCCAACATTTTTAAGGAAAAGTAGAGG 594732
Qy    185 AATAGTTATAGTGTACCCCCATTTACCACCATCACTCAGTTTCAACAGCACATATTTA 244
          |||| ||| ||||   ||||  ||||||| ||||||||||| |||||||||||||||||
Db 594731 ATAAAGAAAACCAGTGCTCCTGCAAAAATGCCGATTTCAGTCGGGGTG-GAGAGAAA 594673
Qy    245 TTTCTTCTATACCAGTACCGTACTCTCCCCACTGGGGATTATATTTAAGGCAAAACCCAGAT 304
          |||||| || ||||| ||||||||| |||||||||| |||||||| |||||| ||||||
Db 594672 TACAAGATCAGCCTCGGAGCATCCGTAGTGACAGAAAAGTAAGGAAAATGCCTAAAACCAAT 594613
Qy    305 G------------ACATTTTTATCCCTAAATACTTTAGATAAAGGTGTTC 341
          |             |||| || || |||| |||||||||||||||||||
Db 594612 GATGGAGGCTAGGCACACGGTGGCTCACACCTGTAATCCCAGCACTTTGGGAGGCCGACAGG 594553
Qy    342 TTTGAAAAA-----AATCATAACCTCAGGACCTGGCCAACATGGTGAAACCCTGTC 396
          ||||||||      |||| |||| |||||||||||||||||||||||||||||||
Db 594552 CATGGATCACGTGGGGTCAGGAGTTTCAGACCAGCCTGGCCAACATGGTGAAACCCCGTC 594493
Qy    397 TGTACTAAAAATAC-AAAAATTAGCTTGGCATGGTGGCGCACCTGTAATCCCAGCTAC 455
          ||||||||||||||  ||||||||||||||||||||||||||||||||||||||||||
Db ... TGTACTAAAAATAC-AAAAATTAGCTTGGCATGGTGGCGCACCTGTAATCCCAGCTAC
```

```
Db  594492 TCTACTAAAAATACAAAAAATTAGCTGGGTGGTGTGGGGGGGCACCTGTAATCCCAGCTAC 594433
            ||||||||| |  ||||||  |||| | ||| | | |||||||||||||||||||||||||
Qy     456 TCAGGAAGCTGAGGCAGGAGGAGAATCACTTGAATCCGGGAGGCAGGAGGTTGCAGTGAGCTGA 515

Db  594432 TCAGGAGGCTGAGGCAGGAGAATTGCTTGAACCTGGGAGGCAGAGGTTGCAGTGAGCCGA 594373
            |||||||||||||||||||| || |||| ||||||||||||||||| |||||||||||||
Qy     516 GAT--TGCAGTCGAGCCTGGGGGACAGAGCAGAAATGAAAACTTCGTCTCTCAAAAACAAAC 573

Db  594372 GATCATGCCATTGCACTCCAGCCCTGGGCAACAAGGAGCCGAAACTTCCATCTCAAAAAAAAAA 594313
            |||  ||||||||||||  ||||||  |  | |||||| ||||  | ||||||||||||||||
Qy     574 AAACAAAAAAACCACTATACATAAAAATGAACAAATGATGGCCACAATAGCACCAGAGA 630

Db  594312 AAAAAAAAAAACCAAACAAAAAAAAGCAAATGATTGCAGCCATTGTCATAAGATA 594256
```

RESULT 12
; US-09-947-911-86/c
; Sequence 86, Application US/0947911
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig
; TITLE OF INVENTION: COLLECTION OF SINGLE NUCLEOTIDE POLYMORPHISM (SNPs) LOCATED ON CH
; TITLE OF INVENTION: X, METHODS OF DETECTION, AND USES THEREOF
; FILE REFERENCE: CL001300
; CURRENT APPLICATION NUMBER: US/09/947,911
; CURRENT FILING DATE: 2001-09-07
; NUMBER OF SEQ ID NOS: 368
; SEQ ID NO 86
; LENGTH: 1059516
; TYPE: DNA
; ORGANISM: HUMAN
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)..(1059516)
; OTHER INFORMATION: n = A,T,C or G
; US-09-947-911-86
```

Query Match   19.9%;   Score 198.8; DB 5;   Length 1059516;
Best Local Similarity   58.8%; Pred. No. 1.8e-21;
Matches 396; Conservative   0; Mismatches 252; Indels 26; Gaps 2;

Qy   4 GCCACCATGCCTGGCTAATTTCTTAATTTTTGTAGTAGAGACGAGGTTTGCCATGTTGACC 63
          |||||||||||||| |||||||| | |||| ||| |||||||||||||| || ||| |||
Db 760960 GCCACCATGCCTGGCTAATTTTTGTATTTTTAGTAAAGACGAGGTTTCACTATGTTGGCC 760901

Qy   64 AGGCTGGTCTCGAACTCTTGACCTCAGGTGATCCGCCCTCAGCCTCCCAAAAGTGCTG 123
          |||||||| |||| ||||||||||||| || | ||||| |||||| ||||||||||
Db 760900 AGGCTGGTCTCAAAACTTCTGACCTCATGTGATCCACCTGCCTCGGCTTCCCAAAGTGCTG 760841

Qy   124 GGATTATAGGCGTGAGCGCCGCGCCCAGCCAACACATTTTTTAAAT-------------- 168
          |||||| |||| | |||||||||||| || ||||| || |||||
Db 760840 GGATTACAGGCATGAGCCACCACGCCTGGCCTGGCTTTACTATCTTTTGTTACTTCCTTT 760781

Qy   169 ------ACTGAAAAGTAGAGGGAATAGTATTAGTGTACCCCATTTACCCATCACTCA 219
                ||| | |||| ||||||| ||| ||||||| || ||||||||||||||||||
Db 760780 CTGACATCCTTGAAAGGAATACAGAAACAAATTAATTACTCAGATGTTACTTAGTTATG 760721

Qy   220 GTTTCACAGCTCGGTGACATATTTTATTTCTTCTATACCAGTACGTACCTCTCCCACTGG 279
          | | | ||||| ||||| ||||||||||||||||||||||||||||||||||||||
Db 760720 CTAAAAATAACTGTCATAGTATCCTAATAATGGCGATCAACCAATGAACTTTAAATTGCAA 760661

Qy   280 GATTATTTTAAGGCAAAACCCAGATGACATTTTATCCCTAAAATACTTTAGATAAAAGGT-- 337
          ||||| |||| |||| ||||||||| | | || | || |||| ||| || |||||
Db 760660 TTATAGGCTGGGTGCAGTGGCTCATGCCTGTAATCCCACCACTTTGGGAGGCCAAGGTAG 760601

Qy   338 GTTCTTTGAAAAAATCATAACCTCAGGACACAGCCTGCCAACATGGTGAAAACCCTGTCT 397
          | ||| ||||| |||| ||||||||| ||||| ||| ||||||||||||||| |||||||
Db 760600 GTGGATCGCTTCAGGTCAGGAGTTCGAGACCAGCCTGCCCAACATGGTGAAACCCTCATCT 760541

Qy   398 GTACTAAAAATACAAAAAATTAGCTGCTGTGTGGGCACCTGTAATCCCAGCTACTC 457
          ||||| ||||||||||||||||||| ||| || ||||| || |||||||||||||||
Db 760540 CTACTAAGAATACAAAAAATTAGCTGTGGGGCGTTGCACCTGTGGTCCCAGCTACTC 760481

Qy   458 AGGGAAGCTGAGGCAGGAGAATCACTTGAATCCGGGAGGCAGAGGATTGCAGTGAGCTGAGA 517
```

RESULT 13

```
Db  760480 GGGAGGCTGAGGCCTGGGCGACAAGAGAATCACTTGAACCTGGGAGGTGGAGGTTACAGTGAGCCGAGA 760421
            ||| |||||||||  |||||||| | | | || |||||||||||||||| |||
Qy     518 TTGCAGTCGAGCCCTGGGCGACAGAGCAGAAATGAAAACTCTGTCTCAAAAACAAAC 577

Db  760420 TCACACCACTGCATTCCAGCCCTAGGTGAAGGAGTGGAGACTGGTGTCTCAAAAAAAAAAAAC 760361
            || | |||  ||||||||||| | ||  |||| | || ||||| | ||||||||||||||
Qy     578 AAAAAAACCACTATACATAAAAATGAACAAATGATGGCCACAATAGCACCAGAGAATTTTAT 637

Db  760360 AAAAAAACAAAAAAAATACTGCAGTTATATTAGCAATCTTCAGTATTTAAAAATTTTAT 760301
            | ||| |  ||||||||  |  ||| ||||||||
Qy     638 AAATACAGATTCCC 651
           || |||  ||||
Db  760300 GTTATCACCTTCAC 760287
```

; NAME/KEY: CDS
; LOCATION: (93827)..(93944)
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (99913)..(100084)
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (107990)..(108116)
US-10-017-161-2143

Query Match   19.8%;   Score 198;   DB 8;   Length 108317;
Best Local Similarity 64.2%;   Conservative 0;   Mismatches 205;   Indels 16;   Gaps 6;
Matches 397;

```
Qy      3 TGCCACCATGCCTGGCTAATTTCTTATTTTAGTGAGACGAGGTTTTGCCATGTTGAC 62
          ||||||||||||||||||||||| ||||||||  ||||||  |||||||||||||||
Db  19862 TACCACCATGCCCAGCTAATTTT-GTATTTTAGTAGAGAC-AGAGTTCTCCATGTTGGT 19919

Qy     63 CAGGCTGGGTCTCGAACTCTTGACCTCAGGTGATCCGCCTGCCTCAGCCTCCCAAAGTGCT 122
          |||||||||||||||||||||||||||||| ||||||||||||||||||||| |||||||
Db  19920 CAGGCTGGTCTCAAACTCCCAAACCTCAGGTGATCACCCTGCCTCAGCCTCTCAAAGTGCT 19979

Qy    123 GGGATTATAGGCGTGAGCCGCCGCGCCCACCGCCAACACATTTTTTAAAATACTGAAAAGT---- 178
          |||||| || |||||||||||||||||||||||||| ||||| |||||||||||||||
Db  19980 GGGATTACAGGCGTGAGCCACTGCACCTGGGCCTAAGTTAATACATTCTTAAAGATTATTT 20039

Qy    179 -AGAGGGAATAGTTATAGTGTACCCCATTTACCCACACACTCAGTTTCAACAGCTGGTGAC 237
           |||||||||||||||||||||||||||||||||||||||||||||| |||||||||
Db  20040 CACAGGAAAAGCAATACAGTCACAATGATGAAGCAGACAGATTTTACATATTCTGTCCATTACC 20099

Qy    238 ATATTTATTTCTTCTATACCAGTACCGTACTCTCCCCACTGGGATTATTTAAGGCAAAA 297
          |||||||||| ||||||||||||| |||||||| ||||||| |||||| ||||||
Db  20100 TGGGAAAAAGTTCAATATCAGGGAAATCTTCTTTCCATTAAAAAAACAAGCTGGCCAGG 20159

Qy    298 CCCAGATG-----ACATTTTATCCCTAAATACTTTAGTATAAAGGTGTTCTTTTGAAA---- 349
          |||| ||       |||||| |||||||||||||| ||||||||| ||||| |||
Db  20160 CATGGTGGCTCACGCCTGTAATCCTAGCACTTTGGGAGGCTGAGGTGGGTGGGATCACTTG 20219

Qy    350 AAATCATAACCTCAGGACCCAGCCTGCCCACACATGGTGAAACCCGTCTCTGTACTAAAAATA 409
          |||||||||| ||||||||||||| |||||||||||||||||||||||| |||||||||
Db  20220 AGGTCAGGAGTTCAAGACCAGCCTGGCCAATATGTGAAACCCCATCTCTACTAAAAATA 20279

Qy    410 CAAAAAATTAGCTTGGGCCATGGTGTCGGGGCACCTGTAATCCCAGCTACTCAGGAAGCTGAGG 469
          |||||||||||| || |||||||||||||||||||||||||||||||| |||||||||
Db  20280 CAAAAACTAGGCAGGCGTGGTGGCGGGCACCTGTAATCCCAGCTACTCGGGGAGGCTGAGG 20339

Qy    470 CAGGAGAATCACTTGAATCCCGGGAAGCAGAGATTGCAGTGAGCTGAGATTGCAGTCGAGC 529
          |||||||||||||||||||||||||||||||||||||| |||||||| |||| |||||||
Db  20340 CAGGAGAATTGCTTGAACCCAGGAGGCGGAGGTTGCAGTGAGCCGAGATTGCCACCACTGC 20399

Qy    530 CTGGGCGACAGAGCGAGACAGA-AATGAAAACTCTGTCTCAAAAAACAAACAAACAAAAAACCAC 588
          |||||||||||||||||||||  |||||| |||||||||||||||||||||||||||||||
Db  20400 ACTCCAGCCTGGGCAACAGACGAGACTCTGTCTCAAAAAAGAACAAACAAACAAAACAAC 20459

Qy    589 TATACATAAAAAATGAACA 606
          |  | ||| ||| || ||
Db  20460 AACAACAACAAAAAAAAA 20477
```

RESULT 14
US-60-427-982-322
; Sequence 322, Application US/60427982
; GENERAL INFORMATION:
; APPLICANT: Wyeth Research
; APPLICANT: Burczynski, Michael
; APPLICANT: Twine, Natalie
; APPLICANT: Dorner, Andrew
; APPLICANT: Trepicchio, William
; TITLE OF INVENTION: METHODS FOR DIAGNOSING RCC AND/OR SOLID TUMORS
; FILE REFERENCE: AM101055
; CURRENT APPLICATION NUMBER: US/60/427,982
; CURRENT FILING DATE: 2002-11-21
; NUMBER OF SEQ ID NOS: 331

; SOFTWARE: Patentin version 3.1
US-60-427-982-322
; SEQ ID NO 322
; LENGTH: 170245
; TYPE: DNA
; ORGANISM: Homo sapiens

Query Match   19.8%;   Score 198;   DB 9;   Length 170245;
Best Local Similarity 62.9%;   Conservative 0;   Mismatches 205;   Indels 19;   Gaps 4;
Matches 380;

```
Qy      5 CCACCATGCCTGGCTAATTTCTTATTTTAGTGAGACGAGGTTTTGCCATGTTGACCA 64
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  64096 CCACCACCCCTGCCTAATTTTTTGTATTTTAGTAGAGACAGGGTTTCACCATGTTGGCA 64155

Qy     65 GGCTGGGTCTCGAACTCTTGACCTCAGGTGATCCGCCTGCCTCAGCCTCCCAAAGTGCTGG 124
          ||||||||||||||||||||||||||| ||     |||||||||||||||||||||||||
Db  64156 GGATGGGTCTCGAACTCCTGACCTCA--TGAACCGCCCACCTCGGCCTCCCAAAGTGCTGG 64213

Qy    125 GATTATAGGCGTGAGCCGCCGCGCCCACCGCCAACACATTTTTTAAAATACTGAAAAGTAGAGG 184
          ||||| |||||||||||||||| ||||||||||||| ||||||||||||||| |||||||||
Db  64214 GATTACAGACGTGAGCCACCGCGCCCGGCCGAAAATTTCTACAGATCCTTCTTGAGGAGC 64273

Qy    185 AATAGTTATAGTGTACCCCATTTACCCACTCACTCAGTTTCACACAGCTGGTGGACATATTTA 244
          ||||||||||||||||| |||| |||||| |||||||| ||||||||||| ||||||| |||
Db  64274 AAAAAAATAAATCTCTAGCGTGTCTAATTTTTAACTCATTCAGGTTTTTTTTTTTTCTTA 64333

Qy    245 TTTCTTCTATACCAGTACCGTACTCTCCCCACTGGGATTATTTAAGGCAAAACCCAGAT 304
          || ||||| ||||||||||||||||| |||||||| ||||||| ||||||||| |||
Db  64334 GAAGTCCAGGCATTTCTTAATTTTTTTCTTAAAAGAAATTGGCTGGGCGGAAT 64393

Qy    305 GAC-----ATTTTATCCCTAAATACTTTAGTATAAAGGTGTTCTTTTGAAA----AAATCA 355
          ||       |||||||||||||||||||||||||||| ||||| ||| |||||
Db  64394 GGCTCACGCCTGTAATCCCTAACACTTTGGGAGGCTGAGGCGGCAGATAACCTGAGGTCA 64453

Qy    356 TAACCTCAGGACCCAGCCTGGCCCAACACATGGTGAAACCCGTCTCTGTACTAAAAATACAAAAA 415
          |||||||||||||||||||||||||||||||||||||||||||| |||||||||||||||
Db  64454 GGAGTTCAAGACCAGCCTGACCAACATGGTGAAACCCCATCTCTACTAAAAATACAAAAA 64513

Qy    416 TTAGCTTGGGCATGGTGGTGCGTGGGCACCTGTAATCCCAGCTACTCAGGAAGCTGAGGCAGGAG 475
          |||||||||| ||||||||||||| ||||||||||||||||||||||||||| |||||||||
Db  64514 TTAGCCGGGCGTGGTGGTGCGGGCGCCTGTAATCCCAGCTACTCGGGAGGCTGAGGCAGGAG 64573

Qy    476 AATCACTTGAATCCCGGGAAGCAGAGATTGCAGTGAGCTGAG-------ATTGCAGTGA 527
          |||||||| || |||||||||||||||||||||||||||      |||||||||
Db  64574 AATTGCTTGAACCTGGGAGGCGGAGGTTGTAGTGAGCCGAGATCAGTGCCATTGCACTCCA 64633

Qy    528 GCCTGGGCGACAGAGCGAGACAGAAATGAAACTCTGTCTCAAAAAAACAAACAAACAAACCA 587
          ||||||||||||||||||||||||      ||||||||||||||||| || ||||||
Db  64634 GCCTGGGCTACAGAGCGAGACTACATCTCAAAACAAACAAACAAAGCAAATTATTC 64693

Qy    588 CTAT 591
          || |
Db  64694 CTTT 64697
```

RESULT 15
US-09-947-911-286/c
; Sequence 286, Application US/09947911
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig
; TITLE OF INVENTION: COLLECTION OF SINGLE NUCLEOTIDE POLYMORPHISM (SNPs) LOCATED ON
; TITLE OF INVENTION: X, METHODS OF DETECTION, AND USES THEREOF
; FILE REFERENCE: CL001300
; CURRENT APPLICATION NUMBER: US/09/947,911
; CURRENT FILING DATE: 2001-09-07
; NUMBER OF SEQ ID NOS: 368
SEQ ID NO 286
; LENGTH: 3037227
; TYPE: DNA
; ORGANISM: HUMAN
; FEATURE:
; NAME/KEY: misc_feature

```
;    LOCATION: (1)...(3037227)
;    OTHER INFORMATION: n = A,T,C or G
US-09-947-911-286 .

Query Match      19.8%;   Score 197.6;   DB 5;   Length 3037227;
Best Local Similarity   62.8%;   Pred. No. 2e-21;
Matches  374;   Conservative  0;   Mismatches  214;   Indels  8;   Gaps  4;

Qy      5 CCACCATGCCTGGCTAATTTTCTTAATTTTTAGTTAGAGACGAGGGTTTTGCCATGTTGACCA 64
          ||||||  ||| ||||  ||||| | || ||||||  |||| |||| |||||||||| ||||| .
Db 686539 CCACCACGCCCCAGCTAATTTTTGTAATTTTTAGTAGAGADATGGGGGTTTCACCATGTTAGCCA 686480

Qy     65 GGCTGGTCTCGAAACTCTTGACCTCAGGTGATCCGCCTGCCTCCCAAAGTGCTGG 124
          ||||  |||||||||||||||||||||||||| ||||||||||||||||||||||||||
Db 686479 AGCTGTTCTCGGAACTCCTGACCTCAGGCAATCCACCTGCCTCCCAAAGTGCTGG 686420

Qy    125 GATTATAGGCGTGAGCCGCCCGCCCACCCAGCCCAACACATTTTTTAAATACTGAAAAAGTAGAGGG 184
          |||||| || ||||| ||  |||| ||  | |||||| | | |||||||||| || || || |
Db 686419 GATTACAGACGTGAGCCCACAGAGCCTGGCCAATACTTTTTTCAAATGAGTCTTTTCAACCA 686360

Qy    185 AATAGTTATAGTGTACCCCATTTACCCATCACTCAGTTTCAACAGCTGGTGACATATTTA 244
          |  | | | | | ||  | |||| | ||| |||| |   ||||||||  ||
Db 686359 CACACTAGGGGACTAATAAAATGATAACTCGCGTCATAATATTATCACTGCCCATCTTTG 686300

Qy    245 TTTCTTCTTATACCAGTACCGTACTCTCCCCACTGGGATTATTT---TAAGGCAAAAACCCA 301
          ||| ||  || || ||| ||| ||| | ||||||||  ||||| |  | |||||||||||
Db 686299 TCAAATCTTAACATGTTTGCTTCTTTTTAAAGTAATAAAATTGGCCAGGCTCGGTGGCT 686240

Qy    302 GATGACATTTTATCCCTAAATACTTTAGATAAAGGGTGTTC--TTTGAAAAAATCATAAC 359
          ||| |||   || | |||||| | || |  |||| ||| | |  |||||||||||| |||
Db 686239 CATGCCTGTAATCCACCACTCTGGGAGACCAAGGTGGGCAGATTACATGAGTCCAGGAG 686180

Qy    360 CTCAGGACCAGCCTGGCCAACATGGTGAAAACCCTGTCTCTGTACTAAAAAATAC--AAAAATT 417
          |||||||||||||||||||||||||||||||| |||||| || ||||||||||||  |||||||
Db 686179 TTTGAGACCAGCCTGGCCAACATGGTGAAATCCTGTCTCTACTAAAAATACAAAAAAATT 686120

Qy    418 AGCTTGGCATGTGGTGTCGGTGGGCCACCTGTAATCCCAGCTACTCAGGAAGCTGAGGCAGGAGA 477
          |||||||||||||||||| |||| ||||||||||||||| ||||| ||||| |  ||||||||||
Db 686119 AACTGGGTGTGGTGGTGGGCATGCCTGTAATTCCAGCTACTCGGGAGGCTGAGGTGGGAGAA 686060

Qy    478 TCACTTGAATCCGGGAAGCAGAGAGATTGCAGTGAGCTGAGATTGCAGTGCAGCCTGGGCGA 537
          |||| || |  |||| ||||| |||||||| |||||||||||||||||||| || || |||
Db 686059 TCACCTAAGCTTAGGAGGTTGCGTGGCTGCAGTGAGCGGAGATGGCACCACTGCACTCCAGC 686000

Qy    538 CAGAGACAGA-AATGAAAACTCTGTCTCAAAAACAAACAAACAAAAAAAAACCACTATA 592
          ||||||||||  ||  | ||| ||| | || ||| || |||| ||| || |  | |||
Db 685999 CTGGGTGACAGAGCTGAGACTTTGTCTCAAAATAAAATAAAAGTAATAAAAAGCCATA 685944
```

OM nucleic - nucleic search, using sw model

GenCore version 5.1.4_p5_4578
Copyright (c) 1993 - 2003 Compugen Ltd.

Run on: March 29, 2003, 13:11:19 ; Search time 3112.01 Seconds
(without alignments)
9361.127 Million cell updates/sec

Title: US-09-691-219-3_COPY_10000_11000
Perfect score: 1001
Sequence: 1 ttactttaccatcctgggg..........ggcccatgggtgaggacatg 1001

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 2054640 seqs, 1455102878 residues
Total number of hits satisfying chosen parameters:
Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : GenEmbl:*
```
1:  gb_ba:*
2:  gb_htg:*
3:  gb_in:*
4:  gb_om:*
5:  gb_ov:*
6:  gb_pat:*
7:  gb_ph:*
8:  gb_pl:*
9:  gb_pr:*
10: gb_ro:*
11: gb_sts:*
12: gb_sy:*
13: gb_un:*
14: gb_vi:*
15: gb_ba:*
16: em_fun:*
17: em_hum:*
18: em_in:*
19: em_mu:*
20: em_om:*
21: em_or:*
22: em_ov:*
23: em_pat:*
24: em_ph:*
25: em_pl:*
26: em_ro:*
27: em_ro:*
28: em_sts:*
29: em_un:*
30: em_vi:*
31: em_htg_hum:*
32: em_htg_inv:*
33: em_htg_mus:*
34: em_htg_other:*
35: em_htg_pln:*
36: em_htg_rod:*
37: em_htg_mam:*
38: em_htg_vrt:*
39: em_sy:*
40: em_htgo_hum:*
41: em_htgo_mus:*
42: em_htgo_other:*
```

Pred. No. is the number of results predicted by chance to have a

score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.		Score	% Query Match	Length	DB	ID	Description
1		1001	100.0	11101	6	AX411375	AX411375 Sequence
2		985.8	98.5	176006	9	AL391650	AL391650 Human DNA
3		731.6	73.1	2291	9	AK023491	AK023491 Homo sapi
4	c	262.8	26.3	321519	2	AL714004	AL714004 Homo sapi
5	c	261.6	26.1	194871	9	AL118858	AL118858 Human chr
6	c	259.8	26.0	222472	2	AC026091	AC026091 Homo sapi
7		259.8	26.0	229155	2	AC087301	AC087301 Homo sapi
8	c	259.6	25.9	139130	2	AC091504	AC091504 Pan trogl
9		259.6	25.9	173109	9	AC091400	AC091400 Pan trogl
10	c	259.2	25.9	159446	9	AC003669	AC003669 Homo sapi
11		259.2	25.9	167730	2	AC061988	AC061988 Homo sapi
12		257.8	25.8	166424	9	AC011301	AC011301 Homo sapi
13	c	259.2	25.9	199721	9	AC092636	AC092636 Homo sapi
14	c	258.4	25.8	157963	9	AP002986	AP002986 Homo sapi
15		257.8	25.8	170682	2	AC023429	AC023429 Homo sapi
16	c	257.6	25.7	155023	2	AC025574	AC025574 Homo sapi
17		257	25.7	72610	2	AL356693_3	Continuation (4 of
18		257	25.7	120007	9	AF064864	AF064864 Homo sapi
19		257	25.7	177762	2	AL831732	AL831732 Homo sapi
20		257	25.7	340000	9	HS21C083	HS21C083 Homo sapi
21		256.8	25.7	166660	9	AC008073	AC008073 Human DNA
22		256.8	25.7	183382	9	AL157388	AL157388 Human DNA
23	c	256.6	25.7	170071	2	AC084733	AC084733 Homo sapi
24	c	256.4	25.6	259894	9	HUAC002302	HUAC002302 Homo sapi
25		256.4	25.6	261070	2	AC130454	AC130454 Homo sapi
26		256	25.6	4039	9	AB030001	AB030001 Human chr
27	c	256	25.6	44679	2	CH19F21246	CH19F21246 Human DNA
28		256	25.6	130056	9	HS7876810	HS7876810 Human DNA
29	c	256	25.6	140092	2	AL117352	AL117352 Human DNA
30		255.8	25.6	162615	2	AL162615	AL162615 Human DNA
31	c	255.8	25.6	162847	2	AC025980	AC025980 Homo sapi
32	c	255.8	25.6	216024	2	AC093707	AC093707 Pan trogl
33		255.8	25.6	231260	2	AL160172	AL160172 Homo sapi
34	c	255.6	25.6	149397	2	AC004840	AC004840 Homo sapi
35		255.6	25.6	194440	9	AL139316	AL139316 Human chr
36		255.6	25.6	201322	9	CNS01DXH	CNS01DXH Human DNA
37	c	255.6	25.6	202397	2	AC091194	AC091194 Homo sapi
38		255.4	25.5	170041	2	AC074394	AC074394 Homo sapi
39		255.2	25.5	147515	2	AL356423	AL356423 Human DNA
40		255	25.5	99227	2	AC094016	AC094016 Pan trogl
41		255	25.5	121012	9	AC114737	AC114737 Homo sapi
42		254.8	25.5	37391	9	AL137860	AL137860 Human DNA
43		254.8	25.5	118055	9	AC005779	AC005779 Homo sapi
44		254.6	25.5	96625	9	AL365223	AL365223 Human DNA
45		254.4	25.4	129950	9	AC000118	AC000118 Human BAC
						AL031289	AL031289 Human DNA

ALIGNMENTS

RESULT 1
AX411375
```
LOCUS       AX411375               11101 bp    DNA     linear   PAT 14-JUN-2002
DEFINITION  Sequence 3 from Patent WO0224910.
ACCESSION   AX411375
VERSION     AX411375.1  GI:21444025
KEYWORDS    .
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1
  AUTHORS   Wei,M.H., Ketchum,K.A., di Francesco,V. and Beasley,E.M.
  TITLE     Isolated human transporter proteins, nucleic acid molecules
            encoding them, and uses thereof
```

JOURNAL   Patent: WO 0224910-A 3 28-MAR-2002;
           PE Corporation (NY) (US)
FEATURES   Location/Qualifiers
     source   1..11101
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
ORIGIN

BASE COUNT   2403 a   3041 c   3200 g   2448 t   9 others

Query Match
Best Local Similarity   100.0%;   Score 1001;   DB 6;   Length 11101;
Matches 1001; Conservative   0; Mismatches   0; Indels   0; Gaps   0;

Qy   1 TTACTTTTACATTCCTTGGGGACAGGGGAACTCTGTGGCCGGGAGGGCTTGCCTCACTGAGGAGT   60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 10000 TTACTTTTACATTCCTTGGGGACAGGGGAACTCTGTGGCCGGGAGGGCTTGCCTCACTGAGGAGT 10059

Qy   61 CAGGTGGGGGGGCTTCCCAGCCTCCCCAGGGCCCCGGGGCAGTGCGTCAGGTGCGTCTCGGCCC   120
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 10060 CAGGTGGGGGGGCTTCCCAGCCTCCCCAGGGCCCCGGGGCAGTGCGTCAGGTGCGTCTCGGCCC 10119

Qy   121 AAGGTGGGAGGAGGAGGGGGGTTGGGTCAGTGCACTTGCCTCCCACTTTTAAATCTCTGTCTTTCCATC   180
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 10120 AAGGTGGGAGGAGGAGGGGGGTTGGGTCAGTGCACTTGCCTCCCACTTTTAAATCTCTGTCTTTCCATC 10179

Qy   181 TGTGAAATGACCTCTTTGTGCCTTCCCAGCACTGTCATCCTGCCTGTTGTTCTTAGGT   240
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 10180 TGTGAAATGACCTCTTTGTGCCTTCCCAGCACTGTCATCCTGCCTGTTGTTCTTAGGT 10239

Qy   241 AGGTGGGGGTCTTCACGCCCTCCCAGGTGCTGTGAAAAGCACTGGCCTGGAG   300
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 10240 AGGTGGGTGGGTCTTCAGCCCCTCCAGGTCTGTGTGAAAACACTGGCCTGGAG 10299

Qy   301 AGGGGTGGGGGGTTGCTGTGGGGTGGGTGGGGGACTCAACCTC   360
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 10300 AGGGGTGGGGGGGTTGCTGTGGTTGGTCTCATTCCAGGGGACTCAACCTC 10359

Qy   361 CCCTACCCACTACCCCCACCCCCAAGCCATGGCAGGCCCCAGGAACTTGATCCTGG   420
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 10360 CCCTACCCACTACCCCCACCCCCAAGCCATGGCAGGCCCCAGGAACTTGATCCTGG 10419

Qy   421 GCTTTGCCGTATGCCAAGTCCTTACACCCTCTCAAGAGACAGAGTCATTGGCTGGGCACGG   480
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 10420 GCTTTGCCGTATGCCAAGTCCTTACACCCCTCTCAAGAGACAGAGTCATTGGCTGGGCACGG 10479

Qy   481 TGGCTCATGCCTGCAATCCAGCACCTTGGGAGGCCGAGGCAGAGGATGACTTGAGGCC   540
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 10480 TGGCTCATGCCTGCAATCCAGCACCTTGGGAGGCCGAGGCAGGCAGAGGATGACTTGAGGCC 10539

Qy   541 AGGAGTTCGAGACCAGCCTGGCCAATATGGGAAAACCTCATTTCTACTAAAATACAAAA   600
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 10540 AGGAGTTCGAGACCAGCCTGGCCAATATGGGAAAACCTCATTTCTACTAAAAATACAAAA 10599

Qy   601 ACTAACCACTGTGTGTCCTGTGTTTGTTGCCTTGGGATTCCCAGCTACTTCGGGAGGCTGAGGCAGGA   660
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 10600 ACTAACCAGGCCGTGGTGGTCTTGCCTGTTGTCCTGTACTAGCTACTCGGGAGGCTGAGGCAGGA 10659

Qy   661 GAATCGGCTTGAACCCGGGAGGCGGAGGTTGCAGTGAGCTGAGATCACACCACTGCACTCC   720
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 10660 GAATCGGCTTGAACCCGGGAGGCGGAGGTTGCAGTGAGCTGAGATCACACCACTGCACTCC 10719

Qy   721 AGCCTGGGCGACAGAGCGAGACTCCAGCTTAAAAAAAAAAAAAAAAGGAGA   780
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 10720 AGCCTGGGCGACAGAGCGAGACTCCAGCTTAAAAAAAAAAAAAAAAAAGGAGA 10779

Qy   781 CCATCACTCGTGTCCTGCATTCTTACAGATGAAAAAACAGCTCAGAGGTTGAAATCGTTT   840
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 10780 CCATCACTCGTGTCCTGCATTCTTACAGATGAAAAAACAGCTCAGAGGTTGAAATCGTTT 10839

Qy   841 TCCTGAAGTCAGACACCAGCCAGTGCCAGGCAGTCTGGGATTTCTGCCCTCATTTCGGTAGACC   900
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 10840 TCCTGAAGTCAGACACCAGCCAGTGCCAGGCAGTCTGGGATTTCTGCCCTCATTTCGGTAGACC 10899

Qy   901 TTCCTCTACAGCAGGTGCGCCCGGCCCCATGGGGTGAGGGACATG   960
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 10900 TTCCTCTACAGCAGGTGCGCCCGGCCCCATGGGGTGAGGGACATG 10959

Qy   961 CCCCTGGGACCACCAGTGCGCCCGGCCCCATGGGGTGAGGGACATG   1001
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 10960 CCCCTGGGACCACCAGTGCGCCCGGCCCCATGGGGTGAGGGACATG 11000

RESULT 2
AL391650/c
LOCUS   AL391650   176006 bp   DNA   linear   PRI 28-NOV-2001
DEFINITION Human DNA sequence from clone RP11-96L14 on chromosome 1, complete
            sequence.
ACCESSION   AL391650
VERSION   AL391650.18   GI:16944045
KEYWORDS   HTG.
SOURCE   human.
  ORGANISM Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1 (bases 1 to 176006)
  AUTHORS   Heath,P.
  TITLE   Direct Submission
  JOURNAL   Submitted (28-NOV-2001) Wellcome Trust Sanger Institute, Hinxton,
            Cambridgeshire, CB10 1SA, UK. E-mail enquiries:
            humquery@sanger.ac.uk Clone requests: clonerequests@sanger.ac.uk
COMMENT   On Nov 15, 2001 this sequence version replaced gi:16605685.
            During sequence assembly data is compared from overlapping clones.
            Where differences are found these are annotated as variations
            together with a note of the overlapping clone name. Note that the
            variation annotation may not be found in the sequence submission
            corresponding to the overlapping clone, as we submit sequences with
            only a small overlap as described above.
            This sequence was finished as follows unless otherwise noted: all
            regions were either double-stranded or sequenced with an alternate
            chemistry or covered by high quality data (i.e., phred quality >=
            30); an attempt was made to resolve all sequencing problems, such
            as compressions and repeats; all regions were covered by at least
            one plasmid subclone or more than one M13 subclone; and the
            assembly was confirmed by restriction digest. The following
            abbreviations are used to associate primary accession numbers given
            in the feature table with their source databases: Em:, EMBL; Sw:,
            SWISSPROT; Tr:, TREMBL; Wp:, WORMPEP; Information on the WORMPEP
            database can be found at
            http://www.sanger.ac.uk/Projects/C_elegans/wormpep This sequence
            was generated from part of bacterial clone contigs of human
            chromosome 1, constructed by the Sanger Centre Chromosome 1 Mapping
            Group. Further information can be found at
            http://www.sanger.ac.uk/HGP/Chr1
            RP11-96L14 is from the library RPCI-11.1 constructed by the group
            of Pieter de Jong. For further details see
            http://www.chori.org/bacpac/home.htm
            VECTOR: pBACe3.6
            This sequence is the entire insert of clone RP11-96L14 The true
            left end of clone RP11-231P20 is at 72881 in this sequence. The
            true right end of clone RP11-111D20 is at 72886 in this sequence.
FEATURES   Location/Qualifiers
     source   1..176006
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="1"
                     /clone="RP11-96L14"
                     /clone_lib="RPCI-11.1"
     misc_feature   169390..169425
                     /note="Sequence from uni-directional dGTP big dye
                     terminator reads only."
ORIGIN

BASE COUNT   45546 a   43652 c   43353 g   43455 t

Query Match
Best Local Similarity   99.7%;   Score 985.8;   DB 9;   Length 176006;
Matches 998; Conservative   0; Mismatches   2; Indels   1; Gaps   1;

```
RESULT 3
LOCUS       AK023491                2291 bp    mRNA    linear   PRI 01-AUG-2002

Qy          1 TTACTTTTACCATCCTGGGGACAGGGAACTCTGTGGGCCCGAGGCGTGCCCTCACTGAGGAGT 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      23501 TTACTTTTACCATCCTGGGGACAGGGAACTCTGTGGGCCCGAGGCGTGCCCTCACTGAGGAGT 23442

Qy         61 CAGGTGGGGCTTCCCAGCCTCCCCAGGGGCAGTGCTGAGTTGTTGTCTTGACTGTTCTGGCCC 120
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      23441 CAGGTGGGGCTTCCCAGCCTCCCCAGGGGCAGTGCTGAGTTGTTGTCTTGACTGTTCTGGCCC 23382

Qy        121 AAGGTGGGGAGGAGGTGGGTTTGGTCTTCACTGACTGTGCTTGCCTGCCTCCCACTTTAAATTCTGTGTTCTTTCCATC 180
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      23381 AAGGTGGGGTCCTTCACCCCCTTCAGCCCCTCCAGGTCTGTTGAAAAGCACTGGCCTGGAG 23322

Qy        181 TGTGAAATGACCTCTTTGTTGTCCTTCCCAGCCACTGTCATCCTGATCGGCCTGTTGTTCTAGGT 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      23321 TGTGAAATGACCTCTTTGTTGTCCTTCCCAGCCACTGTCATCCTGATCGGCCTGTTGTTCTAGGT 23262

Qy        241 AGGTGGGGTCCTTCAGCCCCTCCAGGTGTTGTGAAAAGTCTGTGGAAAAGCACTGGCCTGGAG 300
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      23261 AGGTGGGGTCCTTCAGCCCCTCCAGGTCTGTTGAAAAGCACTGGCCTGGAG 23202

Qy        301 AGGGGTGGGGGGGTTTGGTCTTCTGTGTGGGGTTGCTGGTCGTGAGCTGAGGACTCAACCTC 360
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      23201 AGGGGTGGGGGGGTTTGGTCTTCTGTGTGGGGTTGCTGGTCGTGAGCTGAGGACTCAACCTC 23142

Qy        361 CCCTACCCAACTACCCCACCCCACCCCAAGCCATGGCAGGCCCCAGGGAACTTGATCCTGG 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      23141 CCCTACCCAACT-CCCCACCCCACCCCAAGCCATGGCAGGCCCCAGGGAACTTGATCCTGG 23083

Qy        421 GCTTTGCCGTATGCCAAGTCCTTACACCCTCTCAAGAGACAGTCATTTGGCTGGGCACGG 480
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      23082 GCTTTGCCGTATGCCAAGTCCTTACACCCTCTCAAGAGACAGTCATTTGGCTGGGCACGG 23023

Qy        481 ACTAACCAGGCGTGTGTGGCTTGGCTTGTGCCTGTGTAATCCCAGCTACTCGGGAGGCAGGA 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      23022 TGGCTCATGCCTGCAATCCCAGCACTTGGGAGGCGGAGGCAGGCGGATCACGAGGTCAGGA 22963

Qy        541 AGGAGTTCGAGACCAGCCTGGCCAATATGGCGAAACCCCATCTCTACTAAAAATACAAA 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      22962 AGGAGTTCGAGACCAGCCTGGCCAATATGGCGAAACCCCATCTCTACTAAAAATACAAA 22903

Qy        601 ACTTTGCCGTATGCCAAGTCCTTACACCCTCTCAAGAGACAGTCATTTGGCTGGGCACGG 660
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      22902 ACTAACCAGGCGTGGTGGCGTTGTGCCTGTAATCCCAGCTACTCGGGAGGCTGAGGCAGGA 22843

Qy        661 GAATCGCTTGAACCGGGAGGCAGCAGGTTGCAGTGAGCTGAGATCACACCACTGCACTCC 720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      22842 GAATCGCTTGAACCGGGAGGCAGCAGGTTGCAGTGAGCCGAGATCACACCACTGCACTCC 22783

Qy        721 AGCCTGGGCGACAGAGCGAGACTCCAGCTCTAAAAAAAAAAAAAAAAAAAGGAGA 780
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      22782 AGCCTGGGCGACAGAGCGAGACTCCAGCTTAAAAAAAAAAAAAAAAAAAGGAGA 22723

Qy        781 CCATCACTCGCTGTCCTGCAATCTTTACAGATGAGAAAAACAGGCTCAGAGGTTGAATCGTTT 840
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      22722 CCATCACTCGCTGTCCTGCAATCTTTACAGATGAGAAAAACAGGCTCAGAGGTTGAATCGTTT 22663

Qy        841 TCCTTGAAGTCAGAGCAGAGCCAGTGCAGGCAGGTCTCGGGATTTTCTGCCTCATTCCGGTAGACC 900
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      22662 TCCTTGAAGTCAGAGCAGAGCCAGTGCAGGCAGGTCTCGGGATTTTCTGCCTCATTCCGGTAGACC 22603

Qy        901 TTCCTCTACAGCAGGGTCTGGGCCGGCTGTCGTGGTCTGCGCTGCGCTGTTGGGTACAAATACAAA 960
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      22602 TTCCTCTACAGCAGGGTCTGGGCCGGCTGTCGTGGTCTGCGCTGCGCTGTTGGGTACAAATACAAA 22543

Qy        961 CCCCTGGGACCAGCAGTGCCGGCCCAGCAGGCGCCGGCCCATGGGTGGGTGAGGACACATG 1001
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      22542 CCCCTGGGACCAGCAGTGCCGGCCCAGCAGGCGCCGGCCCATGGGTGGGTGAGGACACATG 22502
```

```
Qy          1 TTACTTTTACCATCCTGGGGACAGGGAACTCTGTGGGCCCGAGGCGTGCCCTCACTGAGGAGT 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1543 TTACTTTTACCATCCTGGGGACAGGGAACTCTGTGGGCCCGAGGCGTGCCCTCACTGAGGAGT 1602

Qy         61 CAGGTGGGGCTTCCCAGCCTCCCCAGGGGCAGTGCTGAGTTGTTGTCTTGACTGTTCTGGCCC 120
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1603 CAGGTGGGGCTTCCCAGCCTCCCCAGGGGCAGTGCTGAGTTGTTGTCTTGACTGTTCTGGCCC 1662

Qy        121 AAGGTGGGGAGGAGGTGGGTTTGGTCTTCACTGACTGTGCTTGCCTGCCTCCCACTTTAAATTCTGTGTTCTTTCCATC 180
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1663 GAGGTGGGGAGGAGGTGGGTTTGGTCACTTGCTTGCCTGCCTCCCACTTTAAATCTCTGTGTTCTTTCCATC 1722

Qy        181 TGTGAAATGACCTCTTTGTTGTCCTTCCCAGCCACTGTCATCCTGATCGGCCTGTTGTTCTAGGT 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1723 TGTGAAATGACCTCTTTGTTGTCCTTCCCAGCCACTGTCATCCTGATCGGCCTGTTGTTCTAGGT 1782

Qy        241 AGGTGGGGTCCTTCAGCCCCTCCAGGTGTTGTGAAAAGTCTGTGGAAAAGCACTGGCCTGGAG 300
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1783 AGGTGGGGTCCTTCAGCCCCTCCAGGTGTTGTGAAAAGTCTGTGGAAAAGCACTGGCCTGGAG 1842

Qy        301 AGGGGTGGGGGGGTTTGGTCTTCTGTGTGGGGTTGCTGGTCGTGAGCTGAGGACTCAACCTC 360
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1843 AGGGGTGGGGGGGTTTGGTCTTCTGTGTGGGGTTGCTGGTCGTGAGCTGAGGACTCAACCTC 1902

Qy        361 CCCTACCCAACTACCCCACCCCACCCCAAGCCATGGCAGGCCCCAGGGAACTTGATCCTGG 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1903 ...
```

RESULT 4

```
LOCUS       AL714004             321519 bp    DNA     linear   HTG 21-MAY-2002
DEFINITION  Homo sapiens clone XX-CHR_20-17119-3, *** SEQUENCING IN PROGRESS
            ***, 89 unordered pieces.
ACCESSION   AL714004
VERSION     AL714004.35  GI:21104209
KEYWORDS    HTG; HTGS_PHASE1.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 321519)
  AUTHORS   Plumb,B.
  TITLE     Direct Submission
  JOURNAL   Submitted (20-MAY-2002) Wellcome Trust Sanger Institute, Hinxton,
            Cambridgeshire, CB10 1SA, UK. E-mail enquiries:
            humquery@sanger.ac.uk Clone requests: clonerequest@sanger.ac.uk
COMMENT     On May 22, 2002 this sequence version replaced gi:21068608.
            -------------- Genome Center
            Center: Wellcome Trust Sanger Institute
            Center code: SC
            Web site: http://www.sanger.ac.uk
            Contact: humquery@sanger.ac.uk
            -------------- Project Information
            Center project name: Chr_20-17119-3
            -------------- Summary Statistics
            Sequencing vector: plasmid; L08752; 100% of reads
            Chemistry: Dye-terminator; 100% of reads
            Assembly program: XGAP4; version 4.5
            Consensus quality: 254683 bases at least Q40
            Consensus quality: 281730 bases at least Q30
            Consensus quality: 299177 bases at least Q20
            Insert size: 312719; sum-of-contigs
            Quality coverage: 1.63x in Q20 bases; sum-of-contigs
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 89 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *         1     4746: contig of 4746 bp in length
```

Alignment block:

```
Db  1903  CCCTACCCAACT-CCCCACCCCACCCAAAGCCACCATGGCAGGCCCCCAGGAAACTTGATCCTGG  1961
          |||||||||||| |||||||||||||||||||||||||||||||||||||||||||||||||||
Qy   421  GCTTTGCCGTATGCCAAGTCCTTACACCCCTCTCAAGAGACACAGTCATTGGCTGGGCACGG    480
Db  1962  GCTTTGCCGTATGCCAAGTCCTTACACCCCTTCAAAGAGACAGTCATTGGCTGGGCACGG   2021
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy   481  TGGCTCATGCCTGCAATCCCAGCACTTTGGGAGGCTGAGGCAGGCAGGATCACTTGAGGCC   540
Db  2022  TGGCTCATGCCTGCAATCCCAGCACTTTGGGAGGCCGAGGCAGGCGGATGACTTGAGGCC   2081
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy   541  AGGAGTTCGAGACCAGCCTGGCCAATATGGCGAAAACCTCATTTCTACTAAAAAATACAAAA   600
Db  2082  AGGAGTTCGAGACCAGCCTGGCCAATATGGCGGAAAACCTCATTTCTACTAAAAAATACAAAA   2141
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy   601  ACTAACCAGGCGTGGTCGCTTGTTGTCCCCGTAATCCCAGCTACTCGGGAGGCTGAGGCAGGA   660
Db  2142  ACTAGGCCAGGGCGTGGGTGGCTTGTTGCCCTGTAATCCCAGCTACTCGGGAGGCTGAGGCAGGA   2201
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy   661  GAATCGCTTGAACCGGGGCAGGAGGTTGCAGTGAGCTGAGATCACACCACCTGCACTCC   720
Db  2202  GAATCGCTTGAACCGGGGAGGGGAGGAGGGTTGCAGTGACCTGAGATCACACCACTGCACTCC   2261
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy   721  AGCCTGGGCGACAGAGCGCAAGACTCCAGCTT   750
Db  2262  AGCCTGGGCGACAGAGCGAAGACTCCAGCTT   2291
          ||||||||||||||||||||||||||||||
```

Contig/gap list:

```
*      4747     4846: gap of 100 bp
*      4847     7172: contig of 2326 bp in length
*      7173     7272: gap of 100 bp
*      7273     9289: contig of 2017 bp in length
*      9290     9389: gap of 100 bp
*      9390    11807: contig of 2418 bp in length
*     11808    11907: gap of 100 bp
*     11908    13962: contig of 2055 bp in length
*     13963    14062: gap of 100 bp
*     14063    16347: contig of 2285 bp in length
*     16348    16447: gap of 100 bp
*     16448    18532: contig of 2085 bp in length
*     18533    18632: gap of 100 bp
*     18633    20654: contig of 2022 bp in length
*     20655    20754: gap of 100 bp
*     20755    23021: contig of 2267 bp in length
*     23022    23121: gap of 100 bp
*     23122    25134: contig of 2013 bp in length
*     25135    25234: gap of 100 bp
*     25235    28285: contig of 3051 bp in length
*     28286    28385: gap of 100 bp
*     28386    30542: contig of 2157 bp in length
*     30543    30642: gap of 100 bp
*     30643    33433: contig of 2791 bp in length
*     33434    33533: gap of 100 bp
*     33534    35599: contig of 2066 bp in length
*     35600    35699: gap of 100 bp
*     35700    39866: contig of 4167 bp in length
*     39867    39966: gap of 100 bp
*     39967    50799: contig of 10833 bp in length
*     50800    50899: gap of 100 bp
*     50900    55372: contig of 4473 bp in length
*     55373    55472: gap of 100 bp
*     55473    58398: contig of 2926 bp in length
*     58399    58498: gap of 100 bp
*     58499    61834: contig of 3336 bp in length
*     61835    61934: gap of 100 bp
*     61935    72530: contig of 10596 bp in length
*     72531    72630: gap of 100 bp
*     72631    77695: contig of 5065 bp in length
*     77696    77795: gap of 100 bp
*     77796    80590: contig of 2795 bp in length
*     80591    80690: gap of 100 bp
*     80690    85052: contig of 4362 bp in length
*     85053    85152: gap of 100 bp
*     85153    89287: contig of 4135 bp in length
*     89288    89387: gap of 100 bp
*     89388    91654: contig of 2267 bp in length
*     91655    91754: gap of 100 bp
*     91755    94398: contig of 2644 bp in length
*     94399    94498: gap of 100 bp
*     94499    96771: contig of 2273 bp in length
*     96772    96871: gap of 100 bp
*     96872   103634: contig of 6763 bp in length
*    103635   103734: gap of 100 bp
*    103735   107935: contig of 4201 bp in length
*    107936   108035: gap of 100 bp
*    108035   113903: contig of 5868 bp in length
*    113904   114003: gap of 100 bp
*    114004   116208: contig of 2205 bp in length
*    116209   116308: gap of 100 bp
*    116309   118998: contig of 2690 bp in length
*    118999   119098: gap of 100 bp
*    119099   122033: contig of 2935 bp in length
*    122034   122133: gap of 100 bp
*    122134   124238: contig of 2105 bp in length
*    124239   124338: gap of 100 bp
*    124339   126514: contig of 2176 bp in length
*    126515   126614: gap of 100 bp
*    126615   136561: contig of 9947 bp in length
*    136562   136661: gap of 100 bp
*    136662   139137: contig of 2476 bp in length
*    139138   139237: gap of 100 bp
```

```
*   139238  142728: contig of 3491 bp in length
*   142729  142828: gap of 100 bp
*   142829  165779: contig of 22951 bp in length
*   165780  165879: gap of 100 bp
*   165880  168731: contig of 2852 bp in length
*   168732  168831: gap of 100 bp
*   168832  170924: contig of 2093 bp in length
*   170925  171024: gap of 100 bp
*   171025  173244: contig of 2220 bp in length
*   173245  173344: gap of 100 bp
*   173345  176010: contig of 2666 bp in length
*   176011  176110: gap of 100 bp
*   176111  178420: contig of 2310 bp in length
*   178421  178520: gap of 100 bp
*   178521  181160: contig of 2640 bp in length
*   181161  181260: gap of 100 bp
*   181261  184022: contig of 2762 bp in length
*   184023  184122: gap of 100 bp
*   184123  186233: contig of 2111 bp in length
*   186234  186333: gap of 100 bp
*   186334  188474: contig of 2141 bp in length
*   188475  188574: gap of 100 bp
*   188575  190834: contig of 2260 bp in length
*   190835  190934: gap of 100 bp
*   190935  193769: contig of 2835 bp in length
*   193770  193869: gap of 100 bp
*   193870  196119: contig of 2250 bp in length
*   196120  196219: gap of 100 bp
*   196220  200732: contig of 4513 bp in length
*   200733  200832: gap of 100 bp
*   200833  204093: contig of 3261 bp in length
*   204094  204193: gap of 100 bp
*   204194  206520: contig of 2327 bp in length
*   206521  206620: gap of 100 bp
*   206621  208679: contig of 2059 bp in length
*   208680  208779: gap of 100 bp
*   208780  210932: contig of 2153 bp in length
*   210933  211032: gap of 100 bp
*   211033  213133: contig of 2101 bp in length
*   213134  213233: gap of 100 bp
*   213234  217420: contig of 4187 bp in length
*   217421  217520: gap of 100 bp
*   217521  222145: contig of 4625 bp in length
*   222146  222245: gap of 100 bp
*   222246  224782: contig of 2537 bp in length
*   224783  224882: gap of 100 bp
*   224883  227854: contig of 2972 bp in length
*   227855  227954: gap of 100 bp
*   227955  230320: contig of 2366 bp in length
*   230321  230420: gap of 100 bp
*   230421  232905: contig of 2485 bp in length
*   232906  233005: gap of 100 bp
*   233006  236533: contig of 3528 bp in length
*   236534  236633: gap of 100 bp
*   236634  239252: contig of 2619 bp in length
*   239253  239352: gap of 100 bp
*   239353  248789: contig of 9437 bp in length
*   248790  248889: gap of 100 bp
*   248890  250940: contig of 2051 bp in length
*   250941  251040: gap of 100 bp
*   251041  253162: contig of 2122 bp in length
*   253163  253262: gap of 100 bp
*   253263  255411: contig of 2149 bp in length
*   255412  255511: gap of 100 bp
*   255512  258118: contig of 2607 bp in length
*   258119  258218: gap of 100 bp
*   258219  260751: contig of 2533 bp in length
*   260752  260851: gap of 100 bp
*   260852  263142: contig of 2291 bp in length
*   263143  263242: gap of 100 bp
*   263243  265368: contig of 2126 bp in length
*   265369  265468: gap of 100 bp
*   265469  268121: contig of 2653 bp in length
```

```
*   268122  268221: gap of 100 bp
*   268222  270223: contig of 2002 bp in length
*   270224  270323: gap of 100 bp
*   270324  274441: contig of 4118 bp in length
*   274442  274541: gap of 100 bp
*   274542  276791: contig of 2250 bp in length
*   276792  276891: gap of 100 bp
*   276892  280982: contig of 4091 bp in length
*   280983  281082: gap of 100 bp
*   281083  283442: contig of 2360 bp in length
*   283443  283542: gap of 100 bp
```

```
Query Match    26.3%;  Score 262.8; DB 2; Length 321519;
Best Local Similarity  89.8%;  Pred. No. 4.8e-61;
Matches   282;  Conservative   0;  Mismatches  32;  Indels   0;  Gaps   0;

Qy     467 TTGGCTGGGCACGGGTGGCTCATGCCTGCAATCCCAGCACCACCTTGGGAGGCTGAGGCAGGCA 526
           |||||||| | |||||||||||| |  |||| |||||||||  |||||||||| ||||| || |
Db  222842 TCGGCTGGGGTGCAGTGGCTCACACCTGTAATCCCAGCACTTTGGGAGGCCGGAGGCGGGGCG 222901

Qy     527 GATGACTTGAGGCCAGGAGTTCGAGACCAGCCTGGCCAATATGGCGAAACCTCATTTCTA 586
           |||||||||||| ||||||||||||||||||||||||||| |||| |||||| |  ||||
Db  222902 GATCACTTGAGGTCAGGAGTTCGAGACCAGCCTGGCCAACATGGTGAAACCCCGTCTCTA 222961

Qy     587 CTAAAAATACAAAAAAACTAACCAGGCGTGGTGGCGGGCGCCTGTGGTCCCAGCTACTCGGG 646
           |||||||||||||||| | | ||||||||||||||||||||||| | ||||||||||||||
Db  222962 CTAAAAATACAAAAAATTAGCCGGGCGTGGTGGCGGGCGCCTGTAATCCCAGCTACTCGGG 223021

Qy     647 AGGGCTGAGGCAGGAGAATCGCTTGAACCCGGGAGGCGGAGGTTGCAGTGAGCTGAGATCA 706
           ||| ||||||||||||||||||||||||||||||||||||||||||||||||| ||||||
Db  223022 AGGCTGAGGCAGGAGAATCGCTTGAACCCGGGAGGCGGAGGTTGCAGTGAGCCGAGATCA 223081

Qy     707 CACCACTGCACTCCAGCCTGGGCGACAAGAGCGAGACTCCGTCTTAAAAAAAAAAAAA 766
           | ||||||||||||||||||||||||||| || ||||||| ||| |||||||||||||
Db  223082 CGCCACTGCACTCCAGCCTGGGCGACAGAGTGAGACTCCATCTCAAAAAAAAAAAAAA 223141

Qy     767 AAAAAAAAGAGA 780
           ||||||||| | |
Db  223142 AAAAAAAAAAAA 223155


RESULT 5
CNS01DRV/c
LOCUS       CNS01DRV     194871 bp    DNA    linear   PRI 04-SEP-2001
DEFINITION  Human chromosome 14 DNA sequence BAC R-1017G21 of library RPCI-11
            from chromosome 14 of Homo sapiens (Human), complete sequence.
ACCESSION   AL118558
VERSION     AL118558.6  GI:15485141
KEYWORDS    HTG; HTGS_ACTIVEFIN.
SOURCE      human.
ORGANISM    Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 194871)
AUTHORS     Heilig,R., Petit,J.L., Vico,V., Dasilva,C., Robert,C., Wincker,P.,
            Brottier,P., Cattolico,L., Barbe,V., Pelletier,E., Artiguenave,F.,
            Levy,M., Eckenberg,R., Bruls,T., deBerardinis,V., Cruaud,C.,
            Gyapay,G., Saurin,W. and Weissenbach,J.
TITLE       Sequencing of the human chromosome 14
JOURNAL     Unpublished
REFERENCE   2  (bases 1 to 194871)
AUTHORS     Genoscope.
TITLE       Direct Submission
JOURNAL     Submitted (04-SEP-2001) Genoscope - Centre National de Sequencage ;
            BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr
            - Web : www.genoscope.cns.fr)
COMMENT     On Sep 6, 2001 this sequence version replaced gi:14571656.
            --------------  Genome Center
            Center: Genoscope / Centre National de Sequencage
            Center code: GS
            Web site: http://www.genoscope.cns.fr/
            Contact: SeqRef@genoscope.cns.fr
            --------------
```

The following BAC sequence is oriented from the T7 to the SP6 end.
Upstream BAC (overlapping the T7 end) : C-2017C7
Downstream BAC (overlapping the SP6 end) : R-114H15 (AC=AL133223)
----------- Summary Statistics
Assembly program: Phrap; version 2.0
Quality coverage: 9.00x in Q20 bases; sum-of-contigs

Overall quality chart :
Range   : bases
-------- --------
0   :
1 - 9   : 9
10 - 19 : 1
20 - 29 : 1
30 - 39 : 36
40 - 49 : 176
50 - 59 : 2561
50 - 59 : 6668
40 - 49 : 12185
60 - 69 : 30197
70 - 79 : 70225
80 - 89 : 72822
90 - 99 :

Percentage of bases with a quality value >= 40 : 99 %.

```
FEATURES             Location/Qualifiers
     source          1..194871
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="14"
                     /clone="R-1017G21"
                     /clone_lib="RPCI-11"
     STS             5321..53394
                     /note="matching EMBL:D26445
                     RHdb:RH35507
                     dbSTS:STS35660
                     Identified using the e-PCR software (G. Schuler)"
     STS             53216..53395
                     /note="matching EMBL:R42676
                     RHdb:RH8632
                     dbSTS:STS13771
                     Identified using the e-PCR software (G. Schuler)"
     STS             55386..55543
                     /note="matching EMBL:G19911
                     RHdb:RH10711
                     dbSTS:STS21713
                     Identified using the e-PCR software (G. Schuler)"
     STS             76990..77117
                     /note="matching EMBL:AA433914
                     RHdb:RH92864
                     dbSTS:STS65867
                     Identified using the e-PCR software (G. Schuler)"
     STS             92484..9604
                     /note="matching EMBL:H97017
                     RHdb:RH92631
                     dbSTS:STS55634
                     Identified using the e-PCR software (G. Schuler)"
     STS             98049..98194
                     /note="matching EMBL:N46343
                     RHdb:RH45764
                     dbSTS:STS38832
                     Identified using the e-PCR software (G. Schuler)"
     STS             161688..161850
                     /note="matching EMBL:T86293
                     RHdb:RH17951
                     dbSTS:STS13765
                     Identified using the e-PCR software (G. Schuler)"
     STS             178303..178552
                     /note="matching EMBL:T03672
                     RHdb:RH53557
                     dbSTS:STS42866
                     Identified using the e-PCR software (G. Schuler)"
     STS             178326..178431
                     Identified using the e-PCR software (G. Schuler)"
                     /note="matching EMBL:G19930
     STS             178392..178577
                     /note="matching EMBL:T15640
                     RHdb:RH1404
                     dbSTS:STS25978
                     Identified using the e-PCR software (G. Schuler)"
                     RHdb:RH10728
                     dbSTS:STS1440
                     Identified using the e-PCR software (G. Schuler)"
     STS             178505..178605
                     /note="matching EMBL:D20215
                     RHdb:RH44085
                     RHdb:RH1617
                     RHdb:RH146
                     dbSTS:STS22994
                     Identified using the e-PCR software (G. Schuler)"
BASE COUNT    49542 a    46484 c    46814 g    52031 t
ORIGIN
```

RESULT 6
AC026091

LOCUS   AC026091   222472 bp   DNA   linear   HTG 17-NOV-2000
DEFINITION Homo sapiens chromosome 17 clone RP11-661C3, WORKING DRAFT
            SEQUENCE, 30 unordered pieces.
ACCESSION   AC026091
VERSION   AC026091.5 GI:10999042
KEYWORDS   HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE   Homo sapiens.
  ORGANISM Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1 (bases 1 to 222472)
  AUTHORS   Waterston,R.H.
  TITLE   The sequence of Homo sapiens clone
  JOURNAL   Unpublished
REFERENCE   2 (bases 1 to 222472)
  AUTHORS   Waterston,R.H.
  TITLE   Direct Submission
  JOURNAL   Submitted (19-MAR-2000) Genome Sequencing Center, Washington
            University School of Medicine, 4444 Forest Park Parkway, St. Louis,
            MO 63108, USA

Query Match   26.1%; Score 261.6; DB 9; Length 194871;
Best Local Similarity   82.4%;
Matches 300; Conservative 0; Mismatches 64; Indels 0; Gaps 0;

```
Qy    412 TGATCCTGGGCGCTTTGCCGTATGCCAAGTCCTTACACCCCTCTCAAGAGACAGTCATTGGC 471
          || |||||||| ||| | |||  ||  |||                        ||   |||
Db  98670 TGTTCCTGGGGTTCTCTTTTATCTTTGTCTTCTTTCCCTTTTAAAATCCCAAAACATGGGC 98611

Qy    472 TGGGCACGGGTGGCTCATGCCCTGCAATCCAGCACCTTGGGAGGCTGAGGTGAGGGAGATGA 531
          ||||||| | ||||||| |||| ||| ||||||| || ||||||| ||||| ||| ||| |
Db  98610 TGGGCACAGTGGCTCACGCCTGCACTCCAGCACTTTGGGAGGCCGAGGCGGGCGGATCA 98551

Qy    532 CTTGAGGCCAGGAGTTCGAGACCAGCCCTGGCCAATATGGCGAAACCTCATTTCTTACTAAA 591
          | ||||| |||||||||||||||||| |||||||| ||| ||||| | || | ||||| 98491
Db  98550 CCTGAGGTCAGGAGTTCGAGACCAGCCTGGCCAACATGGTGAAACCCCATCTCTACTAAA 98491

Qy    592 AATACAAAAACTAACCAGCGGGTGGTGGCTTGGTGTTGTGCCTGTAATCCCAGCTACTCGGGAGGCT 651
          |||||||| || ||  || ||||||||| |   |  | ||||||||||||||||| ||| |||||||
Db  98490 AATACAAAATTACCTAGCACGTGGTGGCACATGCCTGTAATCCCAGCTACTTAGGAGGCT 98431

Qy    652 GAGGCAGGAGAATCGCTTGAACCGGGGAGGCGGAGGTTGCAGTGAGCTGAGATCACACCA 711
          ||| |||||||||| ||||||||   ||||||||||||||||||||| |||||| | |||
Db  98430 GAGACAGGAGAATCACTTGAACCCCAGGAGGCGGAGGTTGCAGTGAGCCGAGATCGTGCCA 98371

Qy    712 CTGCACTCCAGCCTGGGCGACAGAGCGAGACTCCGACTTTAAAAAAAAAAAAAA 771
          |||||||||||||||||  ||||||| ||||  | |   ||||||||||||||
Db  98370 CTGCACTCCAGCCTGGGTGACAGAGCAAGACGCGCATTCTCAAAAAAAAAAAAA 98311

Qy    772 AAAA 775
          ||||
Db  98310 AAAA 98307
```

COMMENT   On Oct 25, 2000 this sequence version replaced gi:9958312.

            ------------------- Genome Center -------------------
            Center: Washington University Genome Sequencing Center
            Center code: WUGSC
            Web site: http://genome.wustl.edu/gsc/index.shtml
            ------------------ Project Information ------------------
            Center project name: H_NH0661C03
            ----------------------- Summary Statistics -----------------------
            Sequencing vector: M13; 100%
            Sequencing vector: plasmid; 0%
            Chemistry: Dye-primer ET; 100% of reads
            Chemistry: Dye-terminator Big Dye; 0% of reads
            Assembly program: Phrap; version 0.990319
            Consensus quality: 206860 bases at least Q40
            Consensus quality: 212309 bases at least Q30
            Consensus quality: 214824 bases at least Q20
            Insert size: 200000; agarose-fp
            Insert size: 226258; sum-of-contigs
            Quality coverage: 4.24 in Q20 bases; agarose-fp
            Quality coverage: 3.81 in Q20 bases; sum-of-contigs

            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 30 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *   1   1448: contig of 1448 bp in length
            *   1449   1548: gap of unknown length
            *   1549   2697: contig of 1149 bp in length
            *   2698   2797: gap of unknown length
            *   2798   4285: contig of 1488 bp in length
            *   4286   4385: gap of unknown length
            *   4386   6203: contig of 1818 bp in length
            *   6204   6303: gap of unknown length
            *   6304   8118: contig of 1815 bp in length
            *   8119   8218: gap of unknown length
            *   8219   10505: contig of 2287 bp in length
            *   10506   10605: gap of unknown length
            *   10606   12521: contig of 1916 bp in length
            *   12522   12621: gap of unknown length
            *   12622   15172: contig of 2551 bp in length
            *   15173   15272: gap of unknown length
            *   15273   17012: contig of 1740 bp in length
            *   17013   17112: gap of unknown length
            *   17113   19838: contig of 2726 bp in length
            *   19839   19938: gap of unknown length
            *   19939   21814: contig of 1876 bp in length
            *   21815   21914: gap of unknown length
            *   21915   25223: contig of 3309 bp in length
            *   25224   25323: gap of unknown length
            *   25324   28416: contig of 3093 bp in length
            *   28417   28516: gap of unknown length
            *   28517   32350: contig of 3834 bp in length
            *   32351   32450: gap of unknown length
            *   32451   36936: contig of 4486 bp in length
            *   36937   37036: gap of unknown length
            *   37037   42124: contig of 5088 bp in length
            *   42125   42224: gap of unknown length
            *   42225   48530: contig of 6306 bp in length
            *   48531   48630: gap of unknown length
            *   48631   53443: contig of 4813 bp in length
            *   53444   53543: gap of unknown length
            *   53544   62383: contig of 8840 bp in length
            *   62384   62483: gap of unknown length
            *   62484   71994: contig of 9511 bp in length
            *   71995   72094: gap of unknown length
            *   72095   79678: contig of 7584 bp in length
            *   79679   79778: gap of unknown length
            *   79779   89178: contig of 9400 bp in length

            *   89179   89278: gap of unknown length
            *   89279   99442: contig of 10164 bp in length
            *   99443   99542: gap of unknown length
            *   99543   111513: contig of 11971 bp in length
            *   111514   111613: gap of unknown length
            *   111614   127267: contig of 15654 bp in length
            *   127268   127367: gap of unknown length
            *   127368   141013: contig of 13647 bp in length
            *   141014   141114: gap of unknown length
            *   141015   141114: gap of unknown length
            *   141115   158417: contig of 17303 bp in length
            *   158418   158517: gap of unknown length
            *   158518   176929: contig of 18412 bp in length
            *   176930   177029: gap of unknown length
            *   177030   196358: contig of 19329 bp in length
            *   196359   196458: gap of unknown length
            *   196459   222472: contig of 26014 bp in length.

FEATURES   Location/Qualifiers
     source   1..222472
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="17"
                     /clone="RP11-661C3"
     misc_feature   1..1448
                     /note="assembly_name:Contig22"
     misc_feature   1549..2697
                     /note="assembly_name:Contig24"
     misc_feature   2798..4285
                     /note="assembly_name:Contig27"
     misc_feature   4386..6203
                     /note="assembly_name:Contig29"
     misc_feature   6304..8118
                     /note="assembly_name:Contig30"
     misc_feature   8219..10505
                     /note="assembly_name:Contig31"
     misc_feature   10606..12521
                     /note="assembly_name:Contig32"
     misc_feature   12622..15172
                     /note="assembly_name:Contig33"
     misc_feature   15273..17012
                     /note="assembly_name:Contig34"
     misc_feature   17113..19838
                     /note="assembly_name:Contig35"
     misc_feature   19939..21814
                     /note="assembly_name:Contig36"
     misc_feature   21915..25223
                     /note="assembly_name:Contig37"
     misc_feature   25324..28416
                     /note="assembly_name:Contig38"
     misc_feature   28517..32350
                     /note="assembly_name:Contig39"
     misc_feature   32451..36936
                     /note="assembly_name:Contig40"
     misc_feature   37037..42124
                     /note="assembly_name:Contig41"
     misc_feature   42225..48530
                     /note="assembly_name:Contig42"
     misc_feature   48631..53443
                     /note="assembly_name:Contig43"
     misc_feature   53544..62383
                     /note="assembly_name:Contig44"
     misc_feature   62484..71994
                     /note="assembly_name:Contig45"
     misc_feature   72095..79678
                     /note="assembly_name:Contig46"
     misc_feature   79779..89178
                     /note="assembly_name:Contig47"
     misc_feature   89279..99442
                     /note="assembly_name:Contig48"
     misc_feature   99543..111513
                     /note="assembly_name:Contig49"
     misc_feature   111614..127267
                     /note="assembly_name:Contig50"
     misc_feature   127368..141014

```
          misc_feature   141115..158417
                         /note="assembly_name:Contig51"
          misc_feature   158518..176929
                         /note="assembly_name:Contig52"
          misc_feature   176930..
                         /note="assembly_name:Contig53"
                         clone_end:t7
                         vector_side:right"
                         /note="177030..196358
          misc_feature   196459..222472
                         /note="assembly_name:Contig54"
                         /note="assembly_name:Contig55"
BASE COUNT    53274 a   56156 c   55982 g   54142 t   2918 others
ORIGIN
```

```
Query Match        26.0%;  Score 259.8;  DB 2;  Length 222472;
Best Local Similarity   83.8%;  Pred. No. 3.2e-60;
Matches  294;  Conservative  0;  Mismatches  57;  Indels  0;  Gaps  0;

Qy   422  CTTTGCCGTATGCCAAGTCCTTACACCCCTCAAGAGACAGCAGTCATTGGCTGGGCACGGT  481
          ||| | |||||  || ||||  || | |||||||| ||  |||||| |||||||||||||||
Db 151423 CTGGGCTATATCCCCAGGGCTTCTCATGATTTAAAAAATCATCCCCAGGGCTGGGCACGGT 151482

Qy   482  GGCTCATGCCCTGCAATCCCACCACCTTGGGAGGGCTGAGGCAGGAGGATGACTTGAGGCCA  541
          ||||||||  || | ||   |   ||||||||| ||||||||| | ||| | ||||| ||
Db 151483 GGCTCATGCCTGTAATCCTAGCACTTTGGGAGGGCAGGAGGATCACCTGAGGTCA 151542

Qy   542  GGAGTTCGAGACCAGCACCAGCCTGGCCAACATAATGGCCGAAAACCTCATTTCTACTAAAAATACAAAAA  601
          |||||||| |||||||||||| |  |||||||||||||||||| |||| ||| ||| | |||||||||||
Db 151543 GGAGTTCGAGACCAGCCTGGGCAATATGGCAAACCCTGTCTCTACTAAAAGTACAAAAA 151602

Qy   602  CTAACCAGGGCGTGGTGGCGTGTGCCTGTAATCCCAGCTACTCGGGAGGCTGAGGCAGGAG  661
          ||| | ||||| |||| ||||| || |||||||||||||||| |||||||||||||||||||
Db 151603 TTAGGCAGGGTGTGGTGGCGTGGGCCTGTAATCCCAGCTACTCAGGAGGCTGAGGCAGGAG 151662

Qy   662  AATGCGCTTGAACCGGGGAGGCGGAGGTTGCAGTGAGCTGAGATCACACCACTGCACTCCA  721
          |||||| |||||  |||||||||||||||  |||||||| |||||| | |||||||||||||
Db 151663 AATGCGCTTGAACCTGGGAGGCGGAGGTTGCAGTGAGCCGAGATCGGCGCCACTGCACTCCA 151722

Qy   722  GCCTGGGCGACGAGAGCGAGACTCCAGCTTAAAAAAAAAAAAAATACAAAAA  772
          |||||||||||||||||||||||||||||||||||||||||||||||||||
Db 151723 GCCTGGTCAATAGAGCAAGACTCAGTCTCAAAAAAAAAAAAAAAAAAA 151773
```

```
RESULT 7
LOCUS       AC087301
            AC087301           229155 bp    DNA    linear     HTG 13-AUG-2002
DEFINITION  Homo sapiens chromosome 17 clone RP11-661c3 map 17, *** SEQUENCING
            IN PROGRESS ***, 1 ordered piece.
ACCESSION   AC087301
VERSION     AC087301.7  GI:22212999
KEYWORDS    HTG; HTGS_PHASE2; HTGS_FULLTOP; HTGS_ACTIVEFIN.
SOURCE      human.
ORGANISM    Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 229155)
AUTHORS     Birren,B., Nusbaum,C. and Lander,E.
TITLE       Homo sapiens chromosome 17, clone RP11-661C3
JOURNAL     Unpublished
REFERENCE   2  (bases 1 to 229155)
AUTHORS     Birren,B., Linton,L., Nusbaum,C., Lander,E., Allen,N., Anderson,S.,
            Barna,N., Bastien,V., Boguslavkiy,L., Boukhgalter,B., Brown,A.,
            Camarata,J., Chang,J., Chazaro,B., Choepel,Y., Collymore,A.,
            Canarata,J., Cooke,P., DeArellano,K., Dewar,K., Diaz,J.S., Dodge,S.,
            Collymore,A., Cooke,P., DeArellano,K., Choepel,Y., Colangelo,M., Collins,S.,
            Dodge,S., Faro,S., Ferreira,P., FitzHugh,W., Gage,D., Galagan,J.,
            Gardyna,S., Ginde,S., Goyette,M., Graham,L., Grand-Pierre,N.,
            Hagos,B., Heaford,A., Horton,L., Hulme,W., Iliev,I., Johnson,R.,
            Jones,C., Karatas,A., LaRocque,K., Lamazares,R., Landers,T.,
            Lehoczky,J., Levine,R., Liu,G., Maclean,C., Macdonald,P.,
            Marquis,N., Matthews,C., McCarthy,M., McEwan,P., McKernan,K.,
            McSheeters,R., Meldrim,J., Meneus,L., Mihova,T., Mlenga,V.,
            Murphy,T., Naylor,J., Nguyen,C., Norbu,C., Norman,C.H.,
            O'Connor,T., O'Donnell,P., O'Neil,D., Oliver,J., Peterson,K.,
            Phunkhang,P., Pierre,N., Pollara,V., Raymond,C., Retta,R.,
            Rieback,M., Riley,R., Rise,C., Rogov,P., Roman,J., Rosetti,M.,
            Roy,A., Santos,R., Schauer,S., Schupback,R., Seaman,S., Severy,P.,
            Sougnez,C., Spencer,B., Stange-Thomann,N., Stojanovic,N.,
            Strauss,N., Subramanian,A., Talamas,J., Tesfaye,S., Theodore,J.,
            Travers,M., Travis,N., Trigilio,J., Vassiliev,H., Viel,R., Vo,A.,
            Wilson,B., Wu,X., Wyman,D., Ye,W.J., Young,G., Zainoun,J.,
            Zembek,L., Zimmer,A. and Zody,M.
TITLE       Direct Submission
JOURNAL     Submitted (23-DEC-2000) Whitehead Institute/MIT Center for Genome
            Research, 320 Charles Street, Cambridge, MA 02141, USA
REFERENCE   3  (bases 1 to 229155)
AUTHORS     Birren,B., Nusbaum,C., Lander,E., Ali,A., Allen,N., Anderson,S.,
            Camarata,J., Chang,J., Chazaro,B., Bloom,T., Boguslavkiy,L., Boukhgalter,B.,
            Cook,A., Cooke,P., DeArellano,K., Dewar,K., Diaz,J.S., Dodge,S.,
            Faro,S., Ferreira,P., FitzGerald,M., Gage,D., Galagan,J.,
            Gardyna,S., Gord,S., Graham,L., Grand-Pierre,N., Hagos,B.,
            Horton,L., Hulme,W., Iliev,I., Johnson,R., Jones,C., Kamat,A.,
            Karatas,A., Kells,C., Landers,T., Levine,R., Lindblad-Toh,K.,
            Liu,G., Maclean,C., Macdonald,P., Major,C., Matthews,C.,
            McCarthy,M., Meldrim,J., Meneus,L., Mihova,T., Mlenga,V.,
            Murphy,T., Naylor,J., Nguyen,C., Nicol,R., Norbu,C., Norman,C.H.,
            O'Connor,T., O'Donnell,P., O'Neil,D., Oliver,J., Peterson,K.,
            Phunkhang,P., Pierre,N., Raymond,C., Retta,R., Rise,C., Rogov,P.,
            Roman,J., Roy,A., Schauer,S., Schupback,R., Seaman,S., Severy,P.,
            Smith,C., Spencer,B., Stange-Thomann,N., Stojanovic,N., Talamas,J.,
            Tesfaye,S., Theodore,J., Topham,K., Travers,M., Vassiliev,H.,
            Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Young,G., Zainoun,J.,
            Zembek,L., Zimmer,A. and Zody,M.
TITLE       Direct Submission
JOURNAL     Submitted (13-AUG-2002) Whitehead Institute/MIT Center for Genome
            Research, 320 Charles Street, Cambridge, MA 02141, USA
COMMENT     On Aug 13, 2002 this sequence version replaced gi:21702875.
            All repeats were identified using RepeatMasker:
            Smit, A.F.A. & Green, P. (1996-1997)
            http://ftp.genome.washington.edu/RM/RepeatMasker.html
            -----------------------------  Genome Center
               Center: Whitehead Institute/ MIT Center for Genome Research
               Center code: WIBR
               Web site: http://www-seq.wi.mit.edu
               Contact: sequence_submissions@genome.wi.mit.edu
            -----------------------------  Project Information
               Center project name: L12041
               Center clone name: 661_C3
            -----------------------------
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 1 contigs. Gaps between the contigs
            * are represented as runs of N. The order of the pieces
            * is believed to be correct as given, however the sizes
            * of the gaps between them are based on estimates that have
            * provided by the submitter.
            * This sequence will be replaced
            * by the finished sequence as soon as it is available and
            * the accession number will be preserved.
            *     1  229155: contig of 229155 bp in length.
FEATURES             Location/Qualifiers
     source          1..229155
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="17"
                     /map="17"
                     /clone="RP11-661c3"
                     /clone_lib="RPCI-11 Human Male BAC"
BASE COUNT    56225 a   58813 c   58166 g   55951 t
ORIGIN
```

```
Query Match        26.0%;  Score 259.8;  DB 2;  Length 229155;
Best Local Similarity   83.8%;  Pred. No. 3.2e-60;
Matches  294;  Conservative  0;  Mismatches  57;  Indels  0;  Gaps  0;

Qy   422  CTTTGCCGTATGCCAAGTCCTTACACCCCTCAAGAGACAGCAGTCATTGGCTGGGCACGGT  481
          ||| | |||||  || ||||  || | |||||||| ||  |||||| |||||||||||||||
```

```
Db 185221 CTGGGCTATATCCCCAGGGCTCTCATGATTTTAAAAAATCATCCCCAGGCTGTGGGCACGGT 185280
          |||||||||||||||||||||||||||||||| |||||||||||||||||||||||||||||
Qy    482 GGCTCATGCGTCGAATCCCAGCACCTTGGGAGGCCGTGAGGCAGGCAGAGATGACTTGAGGCCA 541
Db 185281 GGCTCATGCGTGTAATCCTAGCACTTTGGGAGGATGAGGCAGGCAGAGATCACCTGAGGTCA 185340
          |||||||||||||||||||||||||||||||||||||||||||||||||| ||||||||||
Qy    542 GGAGTTCAGACCAGCCTGGCCAATATGGCGAAACCTCATTTCTACTAAAAAATACAAAA 601
Db 185341 GGAGTTCGAGACCAGCCTGGCCAATATGGCGAAACCCTGTCTCTACTAAAAGTACAAAAA 185400
          |||||||||||||||||||||||||||||||||||||| |||||| ||||||||||||
Qy    602 CTAACCAGGCGTGGTGGCGTGTGCCTGTAATCCCAGCTACTCGGGAGGCTGAGGCAGGAG 661
Db 185401 TTAGCCAGGTGTGGTGGTGTGGGCCTGTAATCCTCAGCTACTCAGGAGGCTGAGGCAGGAG 185460
          ||| |||||||||||||||||||||||||||||||||||||| | |||||||||||||
Qy    662 AATCGCTTGAACCCGGGAGGCGGAGGTTGCAGTGAGCTGAGATCACACCACTGCACTCCA 721
Db 185461 AATTGCTTGAACCTGGGAGGCAGAGGTTGCAGTGAGCTGAGATCGCCACTGCCACTCCA 185520
          || |||||||||| |||||||| |||||||||||||||||||||| |||||| ||||||
Qy    722 GCCTGGGCGGACACAGCGAGACTCCAGCTTAAAAAAAAAAAAAAAAAAAAA 772
Db 185521 GCCTGGTCAATAGAGCAAGAACTCAGTCTCAAAAAAAAAAAAAAAAAAAAAA 185571
```

```
RESULT 8
LOCUS          AC091504/c
DEFINITION     Pan troglodytes clone RP43-13M21, complete sequence.
ACCESSION      AC091504
VERSION        AC091504.2  GI:16303427
KEYWORDS       HTG.
SOURCE         Pan troglodytes
  ORGANISM     Pan troglodytes
               Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
               Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Pan.
REFERENCE      1  (bases 1 to 139130)
  AUTHORS      Akhter,N., Ayele,K., Beckstrom-Sternberg,S.M., Benjamin,B.,
               Blakesley,R.W., Bouffard,G.G., Breen,K., Brinkley,C., Brooks,S.,
               Dietrich,N.L., Granite,S., Guan,X., Gupta,J., Haghighi,P.,
               Ho,S.-L., Idol,J.R., Karlins,E., Laric,P., Lee-Lin,S.-Q.,
               Legaspi,R., Maduro,Q.L., Maduro,V.B., Masiello,C., Mastrian,S.D.,
               McCloskey,J.C., McDowell,J., Pearson,R., Prasad,A., Shevchenko,Y.,
               Stantripop,S., Thomas,J.W., Thomas,P.J., Touchman,J.W.,
               Tsurgeon,C., Vogt,J.L., Walker,M.A., Wetherby,K.D., Wiggins,L.,
               Zhang,L.-H. and Green,E.D.
  TITLE        NISC Comparative Sequencing Initiative
  JOURNAL      Unpublished
REFERENCE      2  (bases 1 to 139130)
  AUTHORS      Green,E.D.
  TITLE        Direct Submission
  JOURNAL      Submitted (27-APR-2001) NIH Intramural Sequencing Center, 8717
               Grovemont Circle, Gaithersburg, MD 20877, USA
REFERENCE      3  (bases 1 to 139130)
  AUTHORS      Green,E.D.
  TITLE        Direct Submission
  JOURNAL      Submitted (20-OCT-2001) NIH Intramural Sequencing Center, 8717
               Grovemont Circle, Gaithersburg, MD 20877, USA
               On Oct 20, 2001 this sequence version replaced gi:13811902.
COMMENT        -------------- Genome Center
                  Center: NIH Intramural Sequencing Center
                  Center code: NISC
                  Web site: http://www.nisc.nih.gov
                  Contact: nisc_mouse@nhgri.nih.gov
               -------------- Project Information
                  Center project name: aoh
                  Center clone name: 013M21
               --------------
               This sequence was finished as follows unless otherwise noted:
               all regions were double-stranded, sequenced with an
               alternate chemistry, or covered by high quality data
               (i.e., phred quality >= 30); an attempt was made to resolve
               all sequencing problems, such as compressions and repeats;
               all regions were covered by at least one plasmid subclone
               or more than one M13 subclone; and the assembly was confirmed
```

```
               by restriction digest.
               CLONE LENGTH: This sequence represents the entire insert of
               this clone unless otherwise noted. If there are overlapping
               clones, the overlaps are noted in the beginning and end of
               the Features section.
FEATURES             Location/Qualifiers
     source          1..139130
                     /organism="Pan troglodytes"
                     /db_xref="taxon:9598"
                     /clone="RP43-13M21"
                     /clone_lib="RP43"
     misc_feature    <1..2243
                     /note="this sequence is not the entire insert of clone
                     RP43-13M21; clone overlaps with GenBank Accession Number
                     AC091296 (nucleotides 152229-185979), clone RP43-109G19
                     (center project name aoj); this annotated segment
                     represents overlap with nucleotides 183737-185979 of
                     AC091296"
     misc_feature    92052..92126
                     /note="single clone coverage"
     misc_feature    139053..139130
                     /note="single clone coverage"
BASE COUNT     33515 a  36608 c  34840 g  34167 t
ORIGIN
```

```
Query Match     25.9%;  Score 259.6;  DB 9;  Length 139130;
Best Local Similarity   81.4%;  Pred. No. 3.6e-60;
Matches  301;  Conservative  0;  Mismatches  69;  Indels  0;  Gaps  0;
```

```
Qy    404 CAGGAACTTGATCCTGGGCTTGCCGTATGCCAAGTCCTTACACCCCTCTCAAGAGACAG 463
          ||||| ||||||||||||| ||||||||||| |||||||||||| |||||||||||||
Db  66072 CATGGAGTGGGTCGGGGCCTGGCTTAGGGTGCCCCTCCCCGACTCTGCTTGTGAGACAAAA 66013
Qy    464 TCATTGGCTGGGCACGGTGGCGTCATGCCTGGGTAATCCCAGCACCTTGGGAGGCTGAGGCAG 523
          ||||||| ||||| |||||||||||||||||| | ||||||||| |||||||||||||||||
Db  66012 GGACTGGCTGGTGCGCGGTGGCTCACGCCTGTAATCCCAGCACTTTGGGAGGCCGAGGCAG 65953
Qy    524 GCAGATGACTTGAGGCCCAGGAGTTCAGACCAGCCTGGCCAATATGGCGAAACCCTCATTT 583
          || ||| ||||||||| ||||||||||||||||||||||||||||||| ||||| |||
Db  65952 GCGGATCACCTGAGGTCAGGAGTTCGAGACCAGCCTGGCCAACATGGTGAAACCCTGTCT 65893
Qy    584 CTACTAAAAAATACAAAAAACTAACCAGGCGTGGTGGCGTGGTGCCATGGTCGCACTCTGTG 643
          ||| |||||||||||||||||||||||||||||||||||| ||||||||| ||||||||||
Db  65892 CTACTAAAAAATACAAAAATTAGCCGGGCATGGTGGCGTGCACCTGTAATCCCAGCTACTC 65833
Qy    644 GGGAGGCTGAGGCAGGAGAATCGCTTGAACCGGGGAGGCGGAGGTTGCAGTGAGCTGAGA 703
          ||||||||||||||| ||||||||||||||||||| ||||||| ||| |||||||||||
Db  65832 GGGAGGCTGAGGCAGGAGAATCGCTTGAACCTGGGGAGGCAGAGGTTGCAGTGAGCCGAGA 65773
Qy    704 TCACACCACTGCACTCCAGCCTGGGCGGACAGCGAGACTCCATCTTAAAAAAAAAAAA 763
          || | ||||||||| |||||||||| ||||| ||||||| | |||||||||||||||||
Db  65772 TGGGCGCCACTGCACTCCAGCCTGGGTGACAGAGCGAGACTCCATCTCAAAAAAAAAAAAA 65713
Qy    764 AAAAAAAAAA 773
          ||||||||||
Db  65712 AAAAAAAAA 65703
```

```
RESULT 9
LOCUS          AC091400
ACCESSION      AC091400
DEFINITION     Pan troglodytes clone RP43-13P21, WORKING DRAFT SEQUENCE, 8
               unordered pieces.
ACCESSION      AC091400
VERSION        AC091400.2  GI:14717331
KEYWORDS       HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE         Pan troglodytes
  ORGANISM     Pan troglodytes
               Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
               Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Pan.
REFERENCE      1  (bases 1 to 173109)
  AUTHORS      Ayele,K., Beckstrom-Sternberg,S.M., Benjamin,B., Blakesley,R.W.,
```

Locus line: `173109 bp    DNA    linear    HTG  13-JUL-2001`

```
TITLE
JOURNAL     Bouffard,G.G., Breen,K., Brinkley,C., Brooks,S., Dietrich,N.L.,
            Granite,S., Guan,X., Gupta,J., Ho,J.-L., Idol,J.R., Karlins,E.,
            Lee-Lin,S.-Q., Legaspi,R., Lin,M., Maduro,Q.L., Maduro,V.B.,
            Masiello,C., Mastrian,S.D., McCloskey,J.C., McDowell,J.,
            Pearson,R., Prasad,A., Shevchenko,Y., Snyder,B., Stantripop,S.,
            Thomas,J.W., Thomas,P.J., Touchman,J.W., Tsurgeon,C., Vogt,J.L.,
            Walker,M.A., Wetherby,K.D., Zhang,L.-H. and Green,E.D.
            NISC Comparative Sequencing Initiative
            Unpublished
REFERENCE   2  (bases 1 to 173109)
AUTHORS     Green,E.D.
TITLE       Direct Submission
JOURNAL     Submitted (19-APR-2001) NIH Intramural Sequencing Center, 8717
            Grovemont Circle, Gaithersburg, MD 20877, USA
COMMENT     On Jul 13, 2001 this sequence version replaced gi:13677071.
            -------------------------------
            Genome Center

            Center: NIH Intramural Sequencing Center
            Center code: NISC
            Web site: http://www.nisc.nih.gov
            Contact: nisc_mouse@nhgri.nih.gov
            -------------------------------
            Project Information

            Center project name: aog
            Center clone name: 013P21
            -------------------------------
            Summary Statistics

            Sequencing vector: plasmid; n/a; 100% of reads
            Chemistry: Dye-terminator Big Dye; 100% of reads
            Assembly program: Phrap; version 0.990319
            Consensus quality: 169170 bases at least Q40
            Consensus quality: 170155 bases at least Q30
            Consensus quality: 170896 bases at least Q20
            Insert size: 150000; pulse-field-gel
            Insert size: 165000; agarose-fp
            Insert size: 172409; sum-of-contigs
            Quality coverage: 12.12x in Q20 bases; agarose-fp
            Quality coverage: 11.12x in Q20 bases; pulse-field-gel
            Quality coverage: 11.02x in Q20 bases; sum-of-contigs
            Quality coverage: 10.55x in Q20 bases; sum-of-contigs
            -------------------------------
            NOTE: This is a 'working draft' sequence. It currently
            consists of 8 contigs. The true order of the pieces
            is not known and their order in this sequence record is
            arbitrary. Gaps between the contigs are represented as
            runs of N, but the exact sizes of the gaps are unknown.
            This record will be updated with the finished sequence
            as soon as it is available and the accession number will
            be preserved.
            *
            *  1          4000:  contig of 4000 bp in length
            *  4001       4100:  gap of unknown length
            *  4101       15655: contig of 11555 bp in length
            *  15656      15755: gap of unknown length
            *  15756      27459: contig of 11704 bp in length
            *  27460      27559: gap of unknown length
            *  27560      47281: contig of 19722 bp in length
            *  47282      47381: gap of unknown length
            *  47382      70102: contig of 22721 bp in length
            *  70103      70202: gap of unknown length
            *  70203      94878: contig of 24676 bp in length
            *  94879      94978: gap of unknown length
            *  94979      127718: contig of 32740 bp in length
            *  127719     127818: gap of unknown length
            *  127819     173109: contig of 45291 bp in length.
FEATURES             Location/Qualifiers
     source          1..173109
                     /organism="Pan troglodytes"
                     /db_xref="taxon:9598"
                     /clone="RP43-13P21"
                     /clone_lib="RP43"
     misc_feature    1..400
                     /note="assembly_fragment"
     misc_feature    4101..15655
                     /note="assembly_fragment"
     misc_feature    15756..27459
                     /note="assembly_fragment"
                     clone_end:SP6
```

```
     misc_feature    vector_side:right"
                     27560..47281
                     /note="assembly_fragment"
     misc_feature    47382..70102
                     /note="assembly_fragment"
     misc_feature    70203..94878
                     /note="assembly_fragment
                     clone_end:T7
                     vector_side:right"
     misc_feature    94979..127718
                     /note="assembly_fragment"
     misc_feature    127819..173109
                     /note="assembly_fragment"
BASE COUNT    41987 a   44656 c   43845 g   41916 t   705 others
ORIGIN

Query Match   25.9%;  Score 259.6;  DB 2;  Length 173109;
Best Local Similarity   81.4%;  Pred. No. 3.6e-60;
Matches  301;  Conservative  0;  Mismatches  69;  Indels  0;  Gaps  0;

Qy    404  CAGGAAACTTGATCCTGGGCCTTGCCGTATGCCAAGTCCTTACACCCTCTCAAGAGACAG  463
           ||  ||||| ||| | |||  |  |  |   |||||||  |||  | | |||||||||||
Db  72725  CANGGAGTGGGGTCGGGGCCTGGCTTAGGGTCCCTCCCGACTCTGCTTTGAGAAGAAA  72784

Qy    464  TCATTGGCTGGGGCACGGTGGCCTCATGCCTGCCAATCCCACACCTTTGGAGGCTGAGGCAG  523
           | |||| |||| ||  |||  | |||| ||||| |||| || ||||| |||||||||||||
Db  72785  GGACTGGCCTGGTCGCGGGTGGCCTCACGCCCTGTAATCCCAGCACTTTGGGAGTCCGGAGGCAG  72844

Qy    524  GCAGATGACTTGAGGCCCAGGAGTTCGAGACCAGCCTGGCCAATATGGCGAAACCTCATTT  583
           || ||| || ||||  ||||||||||  |||||||||  ||| ||  |  |||| || | |
Db  72845  GCGGATCACCTGAGGTCAGGAGTTCGAGACCAGCCTGGCCAACATGGTGAAACCTCGTCT  72904

Qy    584  CTACTAAAAATACAAAAAACTAACCAGGCGTGGTGGTGGCTTGTGCCTGTAATCCCAGCTACTC  643
           ||||||||||| |||||| ||| |||| ||||||||||| ||   | ||||||||||||||||
Db  72905  CTACTAAAAATACAAAAAATTAGCCGGGCATGGTGGTGGCATGCCTGTAATCCCAGCTACTC  72964

Qy    644  GGGAGGCTGAGGCAGGAGAATCGCTTGAACCGGGAGGCGGAGGTTGCAGTGAGCTGAGA  703
           |||||||||||| ||||||| ||| || ||| |||||||||||| | ||||| ||||
Db  72965  GGGAGGCTGAGGCAGGAGAAATCCGTTGAACCTGGGAGGCGGAGGCAGTGAGCCGAGA  73024

Qy    704  TCACACCACTGCACTCCAGCCTGGGCGACAGAGCGAGACTCCAGCTTAAAAAAAAAA  763
           | ||| |||||||||||||||||| ||||||| ||||||||||| ||||||||||||
Db  73025  TGGCGCCACTGCACTCCAGCCTGGGTGACAGAGGGAGACTCCATCTCAAAAAAAAAA  73084

Qy    764  AAAAAAAA 773
           ||||||||
Db  73085  AAAAAAAAA 73094

RESULT 10

LOCUS       AC003669       159446 bp    DNA    linear   PRI 24-MAR-1998
DEFINITION  Homo sapiens Xp22 BAC GS-594A7 (Genome Systems Human BAC library)
            contains Bmx gene, complete sequence.
ACCESSION   AC003669
VERSION     AC003669.1  GI:2981246
KEYWORDS    HTG.
SOURCE      Homo sapiens
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 159446)
  AUTHORS   Muzny,D., Arenson,A.D., Brundage,E., Carvelli,K., Chen,E., Chen,J.,
            Di,W., Ding,Y., Dugan,S., Durbin,J., Forcum,J., Ganesh,R.,
            Garcia,C., Goodman,M., Gorrell,J.H., Haywood,M., Jackson,L.,
            Jin,S., Kampal,R., Karpathy,S., Leal,B., Li,Y., Liu,W., Logan,O.,
            Lu,J., Ly,T., Martinez,C., Oswal,G., Perez,L., Rashid,N.D.,
            Rowland,K., Savage,L., Scherer,S.S., Shen,H., Timms,K.M., Todd,J.,
            Vo,Q., Worley,K.C., Yu,W., Chinault,C., Nelson,D. and Gibbs,R.A.
  TITLE     Direct Submission
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 159446)
  AUTHORS   Worley,K.C.
```

```
TITLE         Direct Submission
JOURNAL       Submitted (09-DEC-1997) Molecular and Human Genetics, Baylor
              College of Medicine, One Baylor Plaza, Houston, TX 77030, USA
REFERENCE     3  (bases 1 to 159446)
AUTHORS       Worley,K.C.
TITLE         Direct Submission
JOURNAL       Submitted (21-MAR-1998) Human Genome Sequencing Center, Department
              of Molecular and Human Genetics, Baylor College of Medicine, One
              Baylor Plaza, Houston, TX 77030, USA
REFERENCE     4  (bases 1 to 159446)
AUTHORS       Worley,K.C.
TITLE         Direct Submission
JOURNAL       Submitted (24-MAR-1998) Human Genome Sequencing Center, Department
              of Molecular and Human Genetics, Baylor College of Medicine, One
              Baylor Plaza, Houston, TX 77030, USA
COMMENT       On Mar 23, 1998 this sequence version replaced gi:2960508.
              Sequencing is completed to a minimum standard of double strand
              coverage with a minimum of 2 clones and 2 reads with no ambiguities
              or 2 chemistries with a minimum of 2 clones and 3 reads with no
              ambiguities. If the sequence quality does not meet this standard,
              it will be indicated in the annotation.

              The repeat regions shown were identified using RepeatMasker by
              Adrian Smit.

              Sequence similarities were identified using Powerblast by Jinghui
              Zhang.

              Exon/Intron boundaries of identified genes were chosen if there
              were canonical splice junctions that maintained sequence continuity
              across the splice junctions.
FEATURES             Location/Qualifiers
     source          1..159446
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="X"
                     /map="Xp22"
                     /clone="GS-594A7"
                     /clone_lib="Genome Systems Human BAC library"
     repeat_region   complement(44..177)
                     /rpt_family="MIR"
     repeat_region   1684..1837
                     /rpt_family="MIR"
     repeat_region   1900..2004
                     /rpt_family="MIR"
     repeat_region   complement(2262..2350)
                     /rpt_family="MIR"
     repeat_region   complement(3789..3811)
                     /rpt_family="AT_rich"
     repeat_region   complement(3875..4149)
                     /rpt_family="AluJo"
     gene            join(6296..6442,7280..7384,9388..9469,14063..14182,
                     15880..15944,20297..20538,23239..23316,23993..24046,
                     27932..27986,29287..29366,32171..32298,34312..34386,
                     35093..35264,39940..40156,41030..41094,44762..44880,
                     47798..47955,54019..54470)
                     /gene="Bmx for cytoplasmic tyrosine kinase"
                     /note="X83107|HSBMXGENE H.sapiens Bmx mRNA for cytoplasmic
                     tyrosine kinase"
     repeat_region   complement(6462..6488)
                     /rpt_family="AT_rich"
     repeat_region   6670..6877
                     /rpt_family="MER20"
     repeat_region   complement(7629..7716)
                     /rpt_family="L2"
     repeat_region   complement(7876..8016)
                     /rpt_family="L2"
     repeat_region   8345..8468
                     /rpt_family="FLAM_C"
     repeat_region   complement(10303..10345)
                     /rpt_family="(TAAAA)n"
     repeat_region   11754..11913
                     /rpt_family="MIR"
```

```
     repeat_region   complement(12319..12420)
                     /rpt_family="MIR"
     repeat_region   12793..12912
                     /rpt_family="L1MC4"
     repeat_region   12859..13273
                     /rpt_family="L1PA8"
     repeat_region   13301..13637
                     /rpt_family="L1MC4"
     repeat_region   complement(14689..14711)
                     /rpt_family="AT_rich"
     misc_feature    complement(17106..17266)
                     /note="Region: AA065091 zf75d12.s1 Soares pineal gland
                     N3HPG Homo sapiens cDNA clone 382775 3',"
     repeat_region   17296..17544
                     /rpt_family="L1MC1C"
     repeat_region   complement(17551..17973)
                     /rpt_family="MSTB"
     repeat_region   17980..18161
                     /rpt_family="MLT1C"
     repeat_region   18266..18464
                     /rpt_family="MLT1C"
     repeat_region   18491..18601
                     /rpt_family="L1MB7"
     STS             18965..18996
                     /gene="Bmx for cytoplasmic tyrosine kinase"
                     /standard_name="HUMSWX1557, Chr. X, Homo sapiens"
                     /db_xref="dbSTS:13053"
     repeat_region   complement(19004..19374)
                     /rpt_family="(CA)n"
     repeat_region   complement(19375..19856)
                     /rpt_family="L1ME3A"
     repeat_region   complement(19857..20131)
                     /rpt_family="AluSq"
     repeat_region   complement(20132..20207)
                     /rpt_family="L1ME3A"
     repeat_region   complement(22277..22381)
                     /rpt_family="(GGAA)n"
     repeat_region   24908..25080
                     /rpt_family="L1MB3"
     repeat_region   25087..25164
                     /rpt_family="AluY"
     repeat_region   complement(24536..24836)
                     /rpt_family="L1MA2"
     repeat_region   25171..25257
                     /rpt_family="(TA)n"
     repeat_region   25258..25298
                     /rpt_family="AluSq"
     repeat_region   26336..26636
                     /rpt_family="L1MA3"
     repeat_region   complement(27012..27082)
                     /rpt_family="AluJb"
     repeat_region   complement(27709..27831)
                     /rpt_family="MIR"
     repeat_region   complement(28157..28434)
                     /rpt_family="MER5B"
     repeat_region   complement(28503..28557)
                     /rpt_family="MLT1F"
     repeat_region   29612..29661
                     /rpt_family="MLT1G"
     repeat_region   29899..30101
                     /rpt_family="L2"
     repeat_region   30617..30676
                     /rpt_family="MIR"
     repeat_region   complement(31152..31435)
                     /rpt_family="AT_rich"
     repeat_region   32587..32618
                     /rpt_family="L1PA12"
     repeat_region   complement(32698..32743)
                     /rpt_family="AT_rich"
     repeat_region   complement(33119..33163)
                     /rpt_family="AT_rich"
     repeat_region   complement(33574..33627)
                     /rpt_family="L1MC/D"
                     /rpt_family="AT_rich"
```

```
repeat_region   complement(34741..34763)
                /rpt_family="AT_rich"
repeat_region   complement(36073..36192)
                /rpt_family="(GA)n"
repeat_region   complement(36243..36353)
                /rpt_family="MER81"
repeat_region   complement(36503..36697)
                /rpt_family="MIR"
repeat_region   complement(36894..37033)
                /rpt_family="MER45"
repeat_region   complement(37577..37620)
                /rpt_family="AT_rich"
repeat_region   37742..38534
                /rpt_family="AT_rich"
STS             38071..38202
                /gene="Bmx for cytoplasmic tyrosine kinase"
                /standard_name="A002D07, Chr. -", Homo sapiens"
                /db_xref="dbSTS:32826"
repeat_region   39396..39468
                /rpt_family="L1PA8"
STS             40989..41161
                /gene="Bmx for cytoplasmic tyrosine kinase"
                /standard_name="HUMSWX1484, Chr. X, Homo sapiens"
                /db_xref="dbSTS:13016"
repeat_region   complement(41389..41513)
                /rpt_family="L1MA2"
repeat_region   41546..41567
                /rpt_family="AT_rich"
repeat_region   complement(41892..41989)
                /rpt_family="L2"
repeat_region   43013..43057
                /rpt_family="AT_rich"


Query Match     25.9%;  Score 259.2;  DB 9;  Length 159446;
Best Local Similarity   86.9%;
Matches  285;  Conservative   0;  Mismatches  43;  Indels   0;  Gaps   0;
                        Pred. No. 4.6e-60;

Qy     461 CAGTCATTGGCTGGGCCACGGTGGCTCAAGCCCTGCCAATCCCAGCACCTTGGGAGGCTGAGG 520
           ||| || |||| |||||||||||||||| ||  |||||||||||||| | ||||||||||||||
Db  113767 CAGAAATAGGCCAGGCACGGTGGCTCACACCTGTAATCCCAGCACTTTGGGAGGCTGAGG 113826

Qy     521 CAGGCAGAATGACTTGAGGCCAGGAGTTCGAGACCAGCCTGGCCAATATGGCGAAAACCTCA 580
           |||||||   ||||||||| |||||||||| |||||||||||||||||||| ||||| ||| |
Db  113827 TGGGCAGATCACTTGAGGTCAGGAGTTCGAAACCAGCCTGGCCAATATGGTGAAACCCTG 113886

Qy     581 TTTCTACTAAAAATACAAAAAACTAAACCAGGCGTGGTGGCTTGTGCCTGTAATCCCAGCTA 640
           |||||||||||||||||||||  || ||| |||||||||||| | |||||||||||||||||
Db  113887 TTTCTACTAAAAATACAAAAATTAGCCCGGCGTGGTGGCACATGCCTGTAATCCCAGCTA 113946

Qy     641 CTCGGGAGGCTGAGGCAGGAGAATCGCTTGAACCCGGGAGGCAGAGGTTGCAGTGAGCTG 700
           || ||||||||||||||||||||||||||||||| |||||||||||||||||||||||||
Db  113947 CTTGGGAGGCTGAGGCAGGAGAATCGCTTGAACCTGGGAGGCAGAGGTTGCAGTGAGCTG 114006

Qy     701 AGATCACACCACTGCACTCCAGCCTGGGCGACAGAGCGAGACTCCAGCTTAAAAAAAAA 760
           |||| |||||||||||||||||||||||| |||||| |||||||||| || ||||||||||||
Db  114007 AGATTGCACCACTGCACTCCAGCCTGGGTGACAGAGTGAGACTCCATCTCAAAAAAAAAAA 114066

Qy     761 AAAAAAAAAAAAAAAAGAGACACATCACT 788
           |||||||||  ||| || ||| | ||| ||
Db  114067 AAAAAAAAAGAAAAGAAAAATCAGAAAT 114094


RESULT 11

LOCUS       AC061988               165730 bp    DNA     linear   HTG 25-MAY-2000
DEFINITION  Homo sapiens chromosome 17 clone RP11-630D10 map 17, WORKING DRAFT
            SEQUENCE, 17 unordered pieces.
ACCESSION   AC061988
VERSION     AC061988.2  GI:8045537
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE      Homo sapiens
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 165730)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C. and Lander,E.
  TITLE     Homo sapiens chromosome 17, clone RP11-630D10
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 165730)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C., Lander,E., Abraham,H., Allen,N.,
            Anderson,S., Baldwin,J., Barna,N., Bastien,V., Beda,F.,
            Boguslavskiy,L., Boukhgalter,B., Brown,A., Burkett,G.,
            Campopiano,A., Castle,A., Choepel,Y., Colangelo,N., Collins,S.,
            Collymore,A., Cooke,P., DeArellano,K., Dewar,K., Diaz,J.S.,
            Dodge,S., Domino,M., Doyle,M., Ferreira,P., FitzHugh,W., Gage,D.,
            Galagan,J., Gardyna,S., Ginde,S., Goyette,M., Graham,L.,
            Grand-Pierre,N., Grant,G., Hagos,B., Heaford,A., Horton,L.,
            Howland,J.C., Iliev,I., Johnson,R., Jones,C., Kann,L., Karatas,A.,
            Klein,J., LaRocque,K., Lamazares,R., Landers,T., Lehoczky,J.,
            Levine,R., Lieu,C., Liu,G., Locke,K., Macdonald,P., Marquis,N.,
            McCarthy,M., McEwan,P., McGurk,A., McKernan,K., McPheeters,R.,
            Meldrim,J., Meneus,L., Mihova,T., Miranda,C., Mlenga,V., Morow,J.,
            Murphy,T., Naylor,J., Norman,C.H., O'Connor,T., O'Donnell,P.,
            O'Neil,D., Olivar,T.M., Oliver,J., Peterson,K., Pierre,N.,
            Pisani,C., Pollara,V., Raymond,C., Riley,R., Rogov,P., Rothman,D.,
            Roy,A., Santos,R., Schauer,S., Severy,P., Spencer,B.,
            Stange-Thomann,N., Stojanovic,N., Subramanian,A., Talamas,J.,
            Tesfaye,S., Theodore,J., Tirrell,A., Travers,M., Trigilio,J.,
            Vassiliev,H., Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J.,
            Young,G., Zainoun,J., Zimmer,A. and Zody,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (21-APR-2000) Whitehead Institute/MIT Center for Genome
            Research, 320 Charles Street, Cambridge, MA 02141, USA
COMMENT     On May 25, 2000 this sequence version replaced gi:7630727.
            All repeats were identified using RepeatMasker.
            http://ftp.genome.washington.edu/RM/RepeatMasker.html
            Smit, A.F.A. & Green, P. (1996-1997)
            Center: Whitehead Institute/MIT Center for Genome Research
            Center code: WIBR
            Web site: http://www-seq.wi.mit.edu
            Contact: sequence_submissions@genome.wi.mit.edu
            -------------------- Project Information --------------------
            Center project name: L9745
            Center clone name: 630_D10
            -------------------- Summary Statistics --------------------
            Sequencing vector: M13; M77815; 100% of reads
            Chemistry: Dye-terminator Big Dye; version 0.960731; 100% of reads
            Assembly program: Phrap; version 0.960731
            Consensus quality: 155748 bases at least Q40
            Consensus quality: 160838 bases at least Q30
            Consensus quality: 162861 bases at least Q20
            Insert size: 162000; agarose-fp
            Insert size: 164130; sum-of-contigs
            Quality coverage: 4.7 in Q20 bases; agarose-fp
            Quality coverage: 4.7 in Q20 bases; sum-of-contigs
            --------------------
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 17 contigs. The true order of the pieces is
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *
            *         1    1988: contig of 1988 bp in length
            *      1989    2088: gap of 100 bp
            *      2089    4897: contig of 2809 bp in length
            *      4898    4997: gap of 100 bp
            *      4998    8702: contig of 3705 bp in length
            *      8703    8802: gap of 100 bp
            *      8803   13304: contig of 4502 bp in length
            *     13305   13404: gap of 100 bp
            *     13405   19091: contig of 5687 bp in length
            *     19092   19191: gap of 100 bp
            *     19192   25880: contig of 6689 bp in length
```

```
FEATURES             Location/Qualifiers
     source          1..165730
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="17"
                     /map="17"
                     /clone_lib="RPCI-11 Human Male BAC"
*    25881..25980: gap of 100 bp in length
*    25981..32708: contig of 6728 bp in length
*    32709..32808: gap of 100 bp in length
*    32809..39945: contig of 7137 bp in length
*    39946..40045: gap of 100 bp in length
*    40046..48263: contig of 8218 bp in length
*    48264..48363: gap of 100 bp in length
*    48364..57091: contig of 8728 bp in length
*    57092..57191: gap of 100 bp in length
*    57192..65288: contig of 8097 bp in length
*    65289..65388: gap of 100 bp in length
*    65389..72736: contig of 7348 bp in length
*    72737..72836: gap of 100 bp in length
*    72837..88513: contig of 15677 bp in length
*    88514..88613: gap of 100 bp in length
*    88614..104826: contig of 16213 bp in length
*    104827..104926: gap of 100 bp in length
*    104927..120656: contig of 15730 bp in length
*    120657..120756: gap of 100 bp in length
*    120757..140643: contig of 19887 bp in length
*    140644..140743: gap of 100 bp in length
*    140744..165730: contig of 24987 bp in length.
     misc_feature    1..1988
                     /note="assembly_fragment"
     misc_feature    2089..4897
                     /note="assembly_fragment"
     misc_feature    4998..8702
                     /note="assembly_fragment"
     misc_feature    8803..13304
                     /note="assembly_fragment"
     misc_feature    13405..19091
                     /note="assembly_fragment"
     misc_feature    19192..25880
                     /note="assembly_fragment"
     misc_feature    25981..32708
                     /note="assembly_fragment"
     misc_feature    32809..39945
                     /note="assembly_fragment"
                     /clone_end:SP6
                     /vector_side:left
     misc_feature    40046..48263
                     /note="assembly_fragment"
     misc_feature    48364..57091
                     /note="assembly_fragment"
     misc_feature    57192..65288
                     /note="assembly_fragment"
     misc_feature    65389..72736
                     /note="assembly_fragment"
                     /clone_end:T7
                     /vector_side:left
     misc_feature    72837..88513
                     /note="assembly_fragment"
     misc_feature    88614..104826
                     /note="assembly_fragment"
     misc_feature    104927..120656
                     /note="assembly_fragment"
     misc_feature    120757..140643
                     /note="assembly_fragment"
     misc_feature    140744..165730
                     /note="assembly_fragment"
BASE COUNT     47628 a   34024 c   3340 g   49137 t   1601 others
ORIGIN

Query Match    25.9%;    Score 259.2;   DB 2;   Length 165730;
Best Local Similarity   86.9%;   Pred. No. 4.6e-60;
```

```
RESULT 12
LOCUS       AC011301      166424 bp    DNA    linear    PRI 07-NOV-2001
DEFINITION  Homo sapiens BAC clone RP11-299I20 from 2, complete sequence.
ACCESSION   AC011301
VERSION     AC011301.9  GI:16756343
KEYWORDS    HTG.
SOURCE      Homo sapiens
ORGANISM    Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 166424)
AUTHORS     Sulston,J.E. and Waterston,R.
TITLE       Toward a complete human genome sequence
JOURNAL     Genome Res. 8 (11), 1097-1108 (1998)
MEDLINE     99063792
PUBMED      9847074
REFERENCE   2  (bases 1 to 166424)
AUTHORS     Kang,K., Maupin,R., Hawkins,M. and Spalding,L.
TITLE       The sequence of Homo sapiens BAC clone RP11-299I20
JOURNAL     Unpublished
REFERENCE   3  (bases 1 to 166424)
AUTHORS     Waterston,R.H.
TITLE       Direct Submission
JOURNAL     Submitted (05-OCT-1999) Genome Sequencing Center, Washington
            University School of Medicine, 4444 Forest Park Parkway, St. Louis,
            MO 63108, USA
REFERENCE   4  (bases 1 to 166424)
AUTHORS     Waterston,R.
TITLE       Direct Submission
JOURNAL     Submitted (07-NOV-2001) Department of Genetics, Washington
            University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
COMMENT     On Nov 7, 2001 this sequence version replaced gi:13431133.
            -------------- Genome Center
            Center: Washington University Genome Sequencing Center
            Center code: WUGSC
            Web site: http://genome.wustl.edu/gsc
            Contact: sapiens@watson.wustl.edu
            -------------- Summary Statistics
            Center project name: H_NH0299I20
            --------------
            NOTICE: This sequence may not represent the entire insert of this
            clone.  It may be shorter because we only sequence overlapping
            clone sections once, or longer because we provide a small overlap
            between neighboring data submissions.

Matches 285;  Conservative  0;  Mismatches  43;  Indels  0;  Gaps  0;

Qy     461 CAGTCATTGGCTGGGCACGGTGGCTCATGCCTGCAATCCCAGCACCTTGGGAGGCTGAGG 520
           ||| ||| ||| |||||||||||||||||||  |||||||||||| |||||||||||||||
Db  110374 CAGAAATAGGCCAGGCGCGGTGGCTCACACCTGTAATCCCAGCACACTTTGGGAGGCTGAGG 110433

Qy     521 CAGGCAGATGACTTGAGGCCCAGGAGTTCGAGACCAGCCTGGCCAATATGGCGAAACCTCA 580
           ||||||||| ||||||||| |||||||||||||||||||||||||||||||||||||||
Db  110434 TGGGCAGATCACTTGAGGTCAGGAGTTCGAGACCAGCCTGGCCAATATGGCGAAACCTG 110493

Qy     581 TTTCTACTAAAAATACAAAAAACTAACCAGGGTGGTGGCGTGTGCCTGTAATCCCAGCTA 640
           |||||||||||||||||||||| || || |||| ||||||||  ||||||||||||||||
Db  110494 TTTCTACTAAAAATACAAAAAATTAGCCGGGCGTGGTGGCGCACATGCCTGTAATCCCAGCTA 110553

Qy     641 CTCGGGAGGCTGAGGCAGGAGAATCGCTTGAACCCGGGAGGCGGAGGTTGCAGTGAGCTG 700
           || |||||||||||||||||||||||||||||||| ||||||||||||||||||||||||
Db  110554 CTTGGGAGGCTGAGGCAGGAGAATCGCTTGAACCTGGGAGGCGGAGGTTGCAGTGAGCTG 110613

Qy     701 AGATCACACCACTGCACTCCAGCCTGGGCGACAGAGCGAGACTCCAGCTTAAAAAAAAAA 760
           |||| ||||||||||||||||||||||||||||||| |||||||| || |||||||||||
Db  110614 AGATTGCACCACTGCACTCCAGCCTGGGCGACAGAGTGAGACTCCATCTCAAAAAAAAAA 110673

Qy     761 AAAAAAAAAAAAAAAAGGAGACCATCACT 788
           |||||||||  ||  ||| ||||| ||| |
Db  110674 AAAAAAAAGAAAGAAAGAAAATCAGAAAT 110701
```

This sequence was finished as follows unless otherwise noted:
all regions were double stranded, sequenced with an alternate
chemistry, or covered by high quality data (i.e., phred quality >=
30); an attempt was made to resolve all sequencing problems, such
as compressions and repeats; all regions were covered by sequence
from more than one subclone; and the assembly was confirmed by
restriction digest.

MAPPING INFORMATION:
Mapping Information for this clone was provided by Dr. John D.
McPherson, Department of Genetics, Washington University, St. Louis
MO. For additional information about the map position of this
sequence, see http://genome.wustl.edu/gsc

SOURCE INFORMATION:
The RPCI-11 human BAC library was made from the blood of one male
donor, as described by Osoegawa,K., Woon,P.Y., Zhao,B., Frengen,E.,
Tateno,M., Catanese,J.J. and de Jong,P.J. (1998) An improved
approach for construction of bacterial artificial chromosome
libraries. Genomics 51:1-8. The clone may be obtained either from
Research Genetics, Inc. (http://www.resgen.com) or Pieter de Jong
and coworkers at the Roswell Park Cancer Institute
(http://bacpac.med.buffalo.edu)
VECTOR: pBACe3.6
NEIGHBORING SEQUENCE INFORMATION:
The clone sequenced to the right is AC024637. Actual start of this
clone is at base position 1 of RP11-299I20; actual end is at base
position 166424 of RP11-299I20.

The sequence H_NH0299I20 from base position 138712 to 140784
contains a tandem repeat. The tandem size is believed to be in
agreement with digest information.

The sequence H_NH0299I20 from base position 141597 to 143810
contains a tandem repeat. The tandem size is believed to be in
agreement with digest information, however, the sequence fidelity
cannot be guaranteed.

The sequence H_NH0299I20 contains a variable CT run from base
position 60019 to 60385. The sequence fidelity cannot be guaranteed
and digest information suggests 400 bases may be missing.

FEATURES   Location/Qualifiers
     source   1. .166424
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="2"
                     /map="2"
                     /clone="RP11-299I20"
                     /clone_lib="RPCI-11"
     repeat_region   2385. .4155
                     /rpt_family="L1"
     repeat_region   4480. .4515
                     /rpt_family="AT_rich"
     repeat_region   4793. .5099
                     /rpt_family="Alu"
     repeat_region   6489. .6793
                     /rpt_family="Alu"
     repeat_region   7161. .7231
                     /rpt_family="MER1_type"
     repeat_region   7645. .7939
                     /rpt_family="(CAGA)n"
     repeat_region   7993. .8068
                     /rpt_family="Alu"
     repeat_region   9062. .9124
                     /rpt_family="L1"
     repeat_region   9127. .9300
                     /rpt_family="Alu"
     repeat_region   9485. .9509
                     /rpt_family="Alu"
     repeat_region   9820. .9988
                     /rpt_family="AT_rich"
     repeat_region   11269. .11312
                     /rpt_family="MER1_type"

     repeat_region   11350. .11690
                     /rpt_family="(TG)n"
     repeat_region   11705. .12011
                     /rpt_family="L1"
     repeat_region   12071. .12139
                     /rpt_family="Alu"
     repeat_region   12140. .12387
                     /rpt_family="Alu"
     repeat_region   12388. .12473
                     /rpt_family="Alu"
     repeat_region   12701. .12993
                     /rpt_family="Alu"
     repeat_region   13470. .13847
                     /rpt_family="Alu"
     repeat_region   14138. .14258
                     /rpt_family="L1"
     repeat_region   14259. .14278
                     /rpt_family="Alu"
     repeat_region   14279. .14490
                     /rpt_family="(TAA)n"
     repeat_region   16282. .16303
                     /rpt_family="Alu"
     repeat_region   16461. .16596
                     /rpt_family="AT_rich"
     repeat_region   16999. .17077
                     /rpt_family="Alu"
     repeat_region   19971. .19999
                     /rpt_family="MIR"
     repeat_region   21318. .21627
                     /rpt_family="AT_rich"
     repeat_region   21878. .22231
                     /rpt_family="Alu"
     repeat_region   22893. .23041
                     /rpt_family="L2"
     repeat_region   24191. .24315
                     /rpt_family="MIR"
     repeat_region   26676. .26728
                     /rpt_family="MIR"
     repeat_region   27053. .27085
                     /rpt_family="(CAAA)n"
     repeat_region   27086. .27308
                     /rpt_family="(TTTAA)n"
     repeat_region   27637. .27679
                     /rpt_family="Alu"
     repeat_region   28753. .28831
                     /rpt_family="A-rich"
     repeat_region   28900. .28920
                     /rpt_family="(CATATA)n"
     repeat_region   30610. .30648
                     /rpt_family="AT_rich"
     repeat_region   33900. .34179
                     /rpt_family="A-rich"
     repeat_region   35297. .35579
                     /rpt_family="Alu"
     repeat_region   37076. .37165
                     /rpt_family="Alu"
     repeat_region   38119. .38231
                     /rpt_family="MIR"
     repeat_region   39830. .40189
                     /rpt_family="L1"
     repeat_region   41100. .41264
                     /rpt_family="L2"
     repeat_region   41312. .41613
                     /rpt_family="MER1_type"
     repeat_region   41628. .41697
                     /rpt_family="Alu"
     repeat_region   42000. .42042
                     /rpt_family="(TA)n"
     repeat_region   42070. .42112
                     /rpt_family="(TC)n"
     repeat_region   42894. .43067
                     /rpt_family="(CA)n"
                     /rpt_family="MIR"

```
repeat_region    43263..43561
                 /rpt_family="Alu"
repeat_region    46025..46116
                 /rpt_family="MIR"
repeat_region    48979..49110
                 /rpt_family="L1"
repeat_region    49048..49157
                 /rpt_family="L1"
repeat_region    49903..49534
                 /rpt_family="L1"
repeat_region    /rpt_family="(CAAA)n"
```

Query Match   25.9%;   Score 259.2; DB 9; Length 166424;
Best Local Similarity 85.7%; Pred. No. 4.6e-60;
Matches 288; Conservative 0; Mismatches 48; Indels 0; Gaps 0;

```
Qy    445 CACCCCTCTCAAGAGACAGTCATTGGCTGCTGGGCACGGTGCTGCTCATGCCCTGCAATCCCAGCA 504
          ||| ||| | ||||||||| ||||  |||||| |||| |||||| ||||  | || ||||||||||
Db  78767 CACCACCCAAAAAGAGAAGGCTTTGGGCCCGGCACAGTGGCTCATTGCCTTGTAATCCCAGCA 78826

Qy    505 CCTTGGGAGGCTGAGGCAGGAGGATCGCTTGAGCCAGGAGTTCGAGACCAGCCTGGGCCA 564
          | ||||||||||||||||||||||||  | |||| |||||||||||||||||||||||||
Db  78827 CTTTGGGAGGCTGAGGCAGGAGGATCACCTGAGGTCAGGAGTTCGAGACCAGCCTGGGCCA 78886

Qy    565 ATATGGGCGAAACCTCATTTCTACTAAAAATACAAAAATTAACCAGGCGTGGTGGCTTGTG 624
          | ||||  ||||| | ||| ||||||||||||||||||||| |||||||||||||| || |
Db  78887 ACATGGTGAAACCCCATCTCTACTAAAAATACAAAAATTAGCCAGGCGTGGTGGCATGCA 78946

Qy    625 CCTGTAATCCCAGCTACTCGGGAGGCTGAGGCAGGAGAATCGCTTGAACCTGGGAGGCAG 684
          ||||||  |||||||||||  |||||||||||||||||||| |||||||||||||||| |
Db  78947 CCTGTAGTCCCAGCTACTCAGGAGGCTGAGGCAGGAGAATTGCTTGAACCTGGGAGGCGG 79006

Qy    685 AGGTTGCAGTGAGCTGAGATCACCACCACTGCACTCCAGCCTGGGCGACAAGAGCGAGACTC 744
          | |||||||||||| ||||| |  |||||||||||||||||||||||||| ||||||||||
Db  79007 AAGTTGCAGTGAGCCGAGATCGTGCCACTGCACTCCAGCCTGGGCGACAGAGCGAGACTC 79066

Qy    745 CAGCTTAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAGGAGA 780
          || || ||||||| ||| || |||| |||| ||| ||| |  |
Db  79067 CATCTCAAAAAAACTAAAACAAAAACAAAAAAGAAAAAAAAACA 79102
```

RESULT 13

```
LOCUS       AC092636    199721 bp    DNA    linear    PRI 16-APR-2002
DEFINITION  Homo sapiens BAC clone RP11-353K11 from 2, complete sequence.
ACCESSION   AC092636 AC068117
VERSION     AC092636.3  GI:19807915
KEYWORDS    HTG.
SOURCE      Homo sapiens.
ORGANISM    Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1 (bases 1 to 199721)
  AUTHORS   Sulston,J.E. and Waterston,R.
  TITLE     Toward a complete human genome sequence
  JOURNAL   Genome Res. 8 (11), 1097-1108 (1998)
  MEDLINE   99063792
   PUBMED   9847074
REFERENCE   2 (bases 1 to 199721)
  AUTHORS   Ali,J., Cotton,M. and Spalding,L.
  TITLE     The sequence of Homo sapiens BAC clone RP11-353K11
  JOURNAL   Unpublished (2001)
REFERENCE   3 (bases 1 to 199721)
  AUTHORS   Waterston,R.H.
  TITLE     Direct Submission
  JOURNAL   Submitted (10-SEP-2001) Genome Sequencing Center, Washington
            University School of Medicine, 4444 Forest Park Parkway, St. Louis,
            MO 63108, USA
REFERENCE   4 (bases 1 to 199721)
  AUTHORS   Waterston,R.H.
  TITLE     Direct Submission
  JOURNAL   Submitted (29-MAR-2002) Genome Sequencing Center, Washington
            University School of Medicine, 4444 Forest Park Parkway, St. Louis,
            MO 63108, USA
REFERENCE   5 (bases 1 to 199721)
  AUTHORS   Waterston,R.
  TITLE     Direct Submission
  JOURNAL   Submitted (16-APR-2002) Department of Genetics, Washington
            University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
COMMENT     On Mar 29, 2002 this sequence version replaced gi:15624963.
            --------------- Genome Center
            Center: Washington University Genome Sequencing Center
            Center code: WUGSC
            Web site: http://genome.wustl.edu/gsc
            Contact: sapiens@watson.wustl.edu
            --------------- Summary Statistics
            Center project name: H_NH0353K11
            Drafting Center: WIBR
            ---------------
            NOTICE: This sequence may not represent the entire insert of this
            clone. It may be shorter because we only sequence overlapping
            clone sections once, or longer because we provide a small overlap
            between neighboring data submissions.

            This sequence was finished as follows unless otherwise noted:
            all regions were double stranded, sequenced with an alternate
            chemistry, or covered by high quality data (i.e., phred quality >-
            30); an attempt was made to resolve all sequencing problems, such
            as compressions and repeats; all regions were covered by sequence
            from more than one subclone; and the assembly was confirmed by
            restriction digest.

            MAPPING INFORMATION:
            Mapping information for this clone was provided by Dr. John D.
            McPherson, Department of Genetics, Washington University, St. Louis,
            MO. For additional information about the map position of this
            sequence, see http://genome.wustl.edu/gsc

            SOURCE INFORMATION:
            The RPCI-11 human BAC library was made from the blood of one male
            donor, as described by Osoegawa,K., Woon,P.Y., Zhao,B., Frengen,E.,
            Tateno,M., Catanese,J.J. and de Jong,P.J. (1998) An improved
            approach for construction of bacterial artificial chromosome
            libraries. Genomics 51:1-8. The clone may be obtained either from
            Research Genetics, Inc. (http://www.resgen.com) or Pieter de Jong
            and coworkers at http://www.chori.org
            VECTOR: pBACe3.6

            NEIGHBORING SEQUENCE INFORMATION:
            The clone sequenced to the left is RP11-145M22; the clone sequenced
            to the right is RP11-219H23. Actual start of this clone is at base
            position 1 of RP11-353K11; actual end is at base position 199721 of
            RP11-353K11.

            The region from 147655 to 147703; 147819 to 147875 are only covered
            by PCR products from clone DNA. Single subclone region exist
            between 147951 and 147966. Polymorphisms exist between AC068539 and
            AC092636. Data from AC068539 and AC079395 was used to finish
            AC092636.

            The sequence of AC068117 has been incorporated into AC092636.
FEATURES             Location/Qualifiers
     source          1..199721
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="2"
                     /map="2"
                     /clone="RP11-353K11"
                     /clone_lib="RPCI-11"
     repeat_region   1..237
                     /rpt_family="Alu"
     repeat_region   1066..1248
                     /rpt_family="MIR"
     repeat_region   1256..1558
                     /rpt_family="Alu"
     repeat_region   1559..1829
                     /rpt_family="Alu"
```

```
repeat_region   1830..2142      /rpt_family="L2"
repeat_region   2143..2327      /rpt_family="Alu"
repeat_region   2473..2757      /rpt_family="L2"
repeat_region   2761..3070      /rpt_family="Alu"
repeat_region   3145..3649      /rpt_family="Alu"
repeat_region   3653..3954      /rpt_family="MaLR"
repeat_region   4149..4314      /rpt_family="Alu"
repeat_region   4832..4858      /rpt_family="Alu"
repeat_region   4867..5177      /rpt_family="(TTTG)n"
repeat_region   5178..5212      /rpt_family="Alu"
repeat_region   5213..5469      /rpt_family="(TTTTG)n"
repeat_region   5613..5818      /rpt_family="Alu"
misc_feature    5650..5665      /note="match to EST BF516553 (NID:g11601732)"
misc_feature    5650..5665      /note="match to EST AI539145 (NID:g4453280) tp73b10.x1"
misc_feature    5650..5665      /note="match to EST AI638881 (NID:g4691115) tt08d08.x1"
repeat_region   5819..6116      /rpt_family="Alu"
misc_feature    5819..5833      /note="match to EST BF516553 (NID:g11601732)"
repeat_region   6117..6170      /rpt_family="(TTTTTA)n"
repeat_region   6171..6452      /rpt_family="Alu"
repeat_region   6453..6592      /rpt_family="Alu"
repeat_region   6594..6683      /rpt_family="MIR"
repeat_region   7074..7200      /rpt_family="MIR"
repeat_region   7685..7851      /rpt_family="MIR"
repeat_region   8320..8349      /rpt_family="Alu"
repeat_region   8353..8653      /rpt_family="(TTTA)n"
repeat_region   8661..8695      /rpt_family="Alu"
repeat_region   8703..8994      /rpt_family="AT_rich"
repeat_region   8995..9100      /rpt_family="Alu"
repeat_region   9102..9185      /rpt_family="Alu"
repeat_region   9880..10190     /rpt_family="MIR"
repeat_region   10235..10395    /rpt_family="Alu"
repeat_region   11104..11148    /rpt_family="MIR"
repeat_region   11169..12025    /rpt_family="L2"
repeat_region   12034..12054    /rpt_family="AT_rich"
repeat_region   13292..13369    /rpt_family="Alu"
repeat_region   13360..13516    /rpt_family="MIR"
repeat_region   13618..13638    /rpt_family="L2"
repeat_region                   /rpt_family="(CTGGGG)n"

misc_feature    13845..14717    /note="CpG island (%GC=72.2, o/e=0.80, #CpGs=97)"
repeat_region   14085..14252    /rpt_family="(CGCGG)n"
repeat_region   14257..14300    /rpt_family="GC_rich"
misc_feature    14307..14756    /rpt_family="GC_rich"
misc_feature    14577..14889    /note="match to EST BM147145 (NID:g17166711)"
misc_feature    14711..15383    /note="match to EST BE247154 (NID:g9098906)"
misc_feature    14839..15010    /note="match to EST BE962211 (NID:g11764881)"
misc_feature    14839..15010    /note="similar to Homo sapiens EST AU130216
                                (NID:g10990570)"
misc_feature    14906..15655    /note="similar to Homo sapiens EST BF922309
                                (NID:g12318197)"
misc_feature    14993..15613    /note="similar to Mus musculus EST BI666503
                                (NID:g15580736)"
misc_feature                    /note="similar to Mus musculus EST BB653323"
```

```
RESULT 14
AP002986/c
LOCUS       AP002986     157963 bp    DNA    linear   PRI 08-MAR-2002
DEFINITION  Homo sapiens genomic DNA, chromosome 11q, clone:RP11-121M22,
            complete sequence.
ACCESSION   AP002986
VERSION     AP002986.2  GI:19263031
KEYWORDS    HTG.
SOURCE      Homo sapiens DNA, clone:RP11-121M22.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 157963)
  AUTHORS   Hattori,M., Ishii,K., Toyoda,A., Taylor,T.D., Hong-Seog,P.,
            Fujiyama,A., Yada,T., Totoki,Y., Watanabe,H. and Sakaki,Y.
  TITLE     Homo sapiens genomic DNA
  JOURNAL   Published Only in Database (2000)
REFERENCE   2  (bases 1 to 157963)
  AUTHORS   Hattori,M., Ishii,K., Toyoda,A., Taylor,T.D., Hong-Seog,P.,
            Fujiyama,A., Yada,T., Totoki,Y., Watanabe,H. and Sakaki,Y.
```

```
Query Match        25.8%;    Score 258.4;  DB 9;  Length 199721;
Best Local Similarity   89.9%;    Pred No. 7.6e-60;
Matches 277;  Conservative 0;  Mismatches 31;  Indels 0;  Gaps 0;

Qy       469 GGCTGGGCACGGTGGCTCATGCCTGTAATCCCAGCACACTTTGGGAGGCTGAGGCAGGCAGA 528
             |||||||| |  |||||||| |||||||||||||||| ||||||||||||||||| || |||
Db    102614 GGCTGGGCGTGCAGTGGCGCTTATGCCTGTAATCCCAGCACTTTGGGAGGCTGAGGTGGGCAGA 102673

Qy       529 TGACTTGAGGCCAGGAGTTCGAGACCAGCCTGGCCAATATGGCGAAACCTCATTTCTACT 588
             | ||||| |||| ||||||||||||||||||||||||| ||| ||||||| ||| |||||
Db    102674 TCACTTGGGGTCAGGAGTTCGAGACCAGCCTGGCCAACATGGTGAAACCCTCATCTCTACT 102733

Qy       589 AAAAATACAAAAAACTAACCAGGCGTGGTGGCGCTTGTGCCTGTAATCCCAGCTACTCGGGAG 648
             |||||||||||||| || |||||||||||||| | |||||||||||||||||  |||||||||
Db    102734 AAAAATACAAAAAATTAGCCAGGCGTGGTGGCATGTGCCTGTAATCCCCCTACTCGGGAG 102793

Qy       649 GCTGAGGCAGGAGAATCGCTTGAACCGGGAGGCAGAGGTTGCAGTGAGCTGAGATCACA 708
             |||||||||||||||||||||||||| ||||||||||||||||||||||| ||||||
Db    102794 GCTGAGGCAGGAGAATCGCTTGAACCTGGGAGGCAGAGGTTGCAGTGAGCCGAGATCGCG 102853

Qy       709 CCACTGCACTCCAGCCTGGGCGACAGAGCGAGACTCCAGCTTAAAAAAAAAAAAAAA 768
             |||||||||||||||||||| |||||| | ||||||| || |||||| |||| ||||
Db    102854 CCACTGCACTCCAGCCTGGGACGAGACGGAGACTCCATCTCAAAAAAATAAAAATAAAAA 102913

Qy       769 AAAAAAAG 776
             |||| |||
Db    102914 AAAATAAG 102921
```

TITLE   Direct Submission
JOURNAL   Submitted (01-DEC-2000) Masahira Hattori, The Institute of Physical
          and Chemical Research (RIKEN), Genomic Sciences Center (GSC);
          1-7-22 Suehiro-chou,Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan
          (E-mail:hattori@gsc.riken.go.jp, URL:http://hgp.gsc.riken.go.jp/,
          Tel:81-45-503-9111, Fax:81-45-503-9170)
COMMENT   On Mar 7, 2002 this sequence version replaced gi:11559301.
FEATURES   Location/Qualifiers
     source   1..157963
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="11"
                     /map="11q"
                     /clone="RP11-121M22"
BASE COUNT   39742 a   36722 c   37247 g   44252 t
ORIGIN

Query Match   25.8%;   Score 257.8;   DB 9;   Length 157963;
Best Local Similarity   85.9%;   Pred. No. 1.1e-59;
Matches 286; Conservative   0; Mismatches 47; Indels   0; Gaps   0;

Qy   469 GGCTGGGCCACGGTGGCTCATGCCTGCAATCCCAGCCACCTTGGGAGGCCGAGGCAGGCAGA 528
             |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   81735 GGCTGGGCCACGGTGGCTCATGCCTGTAATCCCAGCACTTTGGGAGGCCGAGGCAGGCAGA 81676

Qy   529 TGACTTGAGGCCAGGAGTTCGAGACCAGCCTGGCCAATATGGGGAAACCTCATTTCTACT 588
             | || ||||||||||||||||||||||||||||||||||| |||||||| ||||| |||
Db   81675 TTACCTGAGGTCAAGAGTTCGAGACCAGCCTGGCCAACATGGTGAAACCCCATCTCTACT 81616

Qy   589 AAAAATACAAAAAACTAACCAGGCGTGGTGGTTGCTTGTCTGCCTGTAATCCCAGCTACTCGGGAG 648
             ||||||||||| || |||||||||||||||||||||||||||||||||||||||||||||||
Db   81615 AAAAATACAAAAAATTAGCCAGGCGTGGTGGCACACGCCTGTAATCCCAGCTACTTGGGAG 81556

Qy   649 GCTGAGGCAGGAGAATGCTTGAACCCGGGAGGCAGAGGTTGCAGTGAGCTGAGATCACA 708
             ||||||||||||||||||||||||| |||||||||||||||||| ||||||||| |||||
Db   81555 GCTGAGGCAGGAGAATTGCTTGACCCAGGAGGTTGCAGTGAGCCGAGATCATG 81496

Qy   709 CCACTGCACTCCAGCCTGGGCGACAGAGCGAGACTCCAGCTTAAAAAAAAAAAAAA 768
             |||||||||||||||||||||||||||||||||||| ||||||||||||||||||
Db   81495 CCACTGCACTCCAGCCTGGGCTGACAGAGTGAGACTCTGCCTCAAAAAAAAAAAGAAA 81436

Qy   769 AAAAAAAGAGACCATCACTGCTGTCCTGCATT 801
             |||| ||| ||||||||||||||| |||||
Db   81435 GAAAAGAAAATTGGATCATTTCACTTCCATT 81403

RESULT 15
LOCUS   AC023429   170682 bp   DNA   linear   HTG 10-APR-2001
DEFINITION Homo sapiens chromosome 11 clone RP11-121M22, WORKING DRAFT
            SEQUENCE, 7 unordered pieces.
ACCESSION   AC023429
VERSION   AC023429.19 GI:13569974
KEYWORDS   HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_FULLTOP.
SOURCE   Homo sapiens
ORGANISM   Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1 (bases 1 to 170682)
AUTHORS   Abola,A.P., Bruno,D., Conn,L., Dela Rosa,M., Faulkner,D.,
            Federspiel,N., Glukhov,S., Hansen,N., Herman,Z.S., Hyman,R.,
            Mao,J., Komp,C., Kottler,S., Lam,B., Marathe,R., Miranda,M.,
            Morehouse,A.J., Nguyen,M., Oefner,P., Palm,C.J., Ramirez,D.,
            Southwick,A.M., Webb,C., Wilhelmy,J., Yu,S. and Davis,R.W.
JOURNAL   Unpublished
REFERENCE   2 (bases 1 to 170682)
AUTHORS   Bruno,D., Conn,L., Dela Rosa,M., Faulkner,D., Federspiel,N.,
            Glukhov,S., Hansen,N., Hyman,R., Mao,J., Marathe,R.,
            Morehouse,A.J., Oefner,P., Palm,C.J., Ramirez,D., Wilhelmy,J.,
            Yu,S. and Davis,R.W.
TITLE   Direct Submission
JOURNAL   Submitted (14-FEB-2000) DNA Sequencing and Technology Center,
            Stanford University, 855 California Avenue, Palo Alto, CA 94304,
            USA
COMMENT   On Apr 10, 2001 this sequence version replaced gi:13562078.
            ----------------------- Genome Center
            Center: Stanford Genome Center
            Center: Stanford DNA Sequencing and Technology Development
            ----------------------- Project Information
            Center code: SDSTDC
            Web site: http://sequence-www.stanford.edu/group/human/
            Contact: hum-info@sequence.stanford.edu
            ----------------------- Project Information
            Center project name: 837
            Center clone name: RP11-121M22
            ----------------------- Summary Statistics
            Sequencing Vector: M13mp18; X02513
            Chemistry: Dye-primer; 12% of reads
            Chemistry: Dye-terminator Big Dye; 86% of reads
            Assembly program: Phrap; version 0.990319
            Consensus quality: 165770 bases at least Q40
            Consensus quality: 167235 bases at least Q30
            Consensus quality: 167918 bases at least Q20
            Insert size: 170082; sum-of-contigs
            Insert size: 172423; agarose-fp
            Quality coverage: 9.0x in Q20 bases; agarose-fp
            Quality coverage: 9.1x in Q20 bases; sum-of-contigs.
          * NOTE: This is a 'working draft' sequence. It currently
          * consists of 7 contigs. The true order of the pieces
          * is not known and their order in this sequence record is
          * arbitrary. Gaps between the contigs are represented as
          * runs of N, but the exact sizes of the gaps are unknown.
          * This record will be updated with the finished sequence
          * as soon as it is available and the accession number will
          * be preserved.
          *
          *   1   1856: contig of 1856 bp in length
          *   1857   1956: gap of unknown length
          *   1957   7065: contig of 5109 bp in length
          *   7066   7165: gap of unknown length
          *   7166   21089: contig of 13924 bp in length
          *   21090   21189: gap of unknown length
          *   21190   34145: contig of 12956 bp in length
          *   34146   34245: gap of unknown length
          *   34246   48562: contig of 14317 bp in length
          *   48563   48662: gap of unknown length
          *   48663 109179: contig of 60517 bp in length
          *   109180 109279: gap of unknown length
          *   109280 170682: contig of 61403 bp in length
FEATURES   Location/Qualifiers
     source   1..170682
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="11"
                     /clone="RP11-121M22"
                     /clone_lib="RPCI human BAC library 11"
     misc_feature   1..1856
                     /note="assembly_name:Contig25"
     misc_feature   1957..7065
                     /note="assembly_name:Contig26"
     misc_feature   7166..21089
                     /note="assembly_name:Contig27"
     misc_feature   21190..34145
                     /note="assembly_name:Contig28"
     misc_feature   34246..48562
                     /note="assembly_name:Contig29"
     misc_feature   48663..109179
                     /note="assembly_name:Contig30"
                     /clone_end:SP6"
     misc_feature   109280..170682
                     /note="assembly_name:Contig31"
                     /clone_end:T7"
BASE COUNT   43947 a   39732 c   39514 g   46884 t   605 others
ORIGIN

Query Match   25.8%;   Score 257.8;   DB 2;   Length 170682;
Best Local Similarity   85.9%;   Pred. No. 1.1e-59;
Matches 286; Conservative   0; Mismatches 47; Indels   0; Gaps   0;

```
Qy    469 GGCTGGGCACGGTGGGCTCATGGCCTGCAATCCCAGCACCTTGGGAGGCTGAGGCAGGCAGA 528
          ||||||||||||||||||||||||||||||||| |||||||||||||||||||||||||||
Db   9745 GGCTGGGCACGGTGGGCTCATGGCCTGTAATCCCAGCACTTTGGGAGGCTGAGGCAGGCAGA 9804

Qy    529 TGACTTGAGGGCCAGGAGTTCGAGACCAGCCTGGCCAATATGGCGAAACCTCAATTTCTACT 588
          | || ||||| | || |||||||| |||||| ||  ||| ||  ||  ||| |||| ||||
Db   9805 TTACCTGAGGTCAAAGAGTTCGAGACCAGGCCCAACACATGGTGAATCCCCATCTCTACT 9864

Qy    589 AAAAATACAAAAAACTAACCAGGCGTGGTGGCGGGCTTTGTGCCTGTAAATCCCAGCTACTCGGGAG 648
          |||||||| | ||||||||||||||| |||||||||| ||||||||||||||||||| | ||||||
Db   9865 AAAAATACAAAAATTAGCCAGGCGTGGTGGGCACACGCCCTGTAATCCCAGCTRACTTGGGAG 9924

Qy    649 GCTGAGGCAGGAGAATCGCTTGAACCGGGGAGGCAGAGGTTGCAGTGAGCTGAGATCACA 708
          |||||||||||||||| ||||| |||||||||||||||||||||||||||||||| |||
Db   9925 GCTGAGGCAGGAGAATTGCTTGAGCCGGGGAGGCAGAGGTTGCAGTGAGCCGGAGATCATG 9984

Qy    709 CCACTGCACTCCAGCCTGGGCGACAGAGCGAGACTCCAGCTTAAAAAAAAAAAAAAAAAAAA 768
          ||||||||||||||||||||| || || |||||||||||| | ||||||||||||| ||||
Db   9985 CCACTGCACTCCAGCCTGGCTGACAGAGTGAGACTCTGCCTCAAAAAAAAAAAAAAGAAA 10044

Qy    769 AAAAAAAGGAGACCATCACTGCTGTCCTGCATT 801
          |||| |  | |||||| || ||| | || ||||
Db  10045 GAAAAGAAAATTGGATCATTTCACTTCTCCATT 10077
```

Search completed: March 29, 2003, 21:53:09
Job time : 6057.01 secs

OM nucleic - nucleic search, using sw model

GenCore version 5.1.4_p5_4578
Copyright (c) 1993 - 2003 Compugen Ltd.

Run on: March 29, 2003, 08:02:28 ; Search time 273.075 Seconds
                                   (without alignments)
                                   8255.046 Million cell updates/sec

Title: US-09-691-219-3_COPY_10000_11000
Perfect score: 1001
Sequence: 1 ttacttttaccatcctgggg........ggcccatgggtggaggacatg 1001

Scoring table: IDENTITY_NUC
                Gapop 10.0 , Gapext 1.0

Searched: 2185239 seqs, 112599159 residues

Total number of hits satisfying chosen parameters: 4370478

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 08
                 Maximum Match 100%
                 Listing first 45 summaries

Database :   N_Geneseq_101002:*
    1: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1980.DAT:*
    2: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1981.DAT:*
    3: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1982.DAT:*
    4: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1983.DAT:*
    5: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1984.DAT:*
    6: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1985.DAT:*
    7: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1986.DAT:*
    8: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1987.DAT:*
    9: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1988.DAT:*
   10: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1989.DAT:*
   11: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1990.DAT:*
   12: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1991.DAT:*
   13: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1992.DAT:*
   14: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1993.DAT:*
   15: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1994.DAT:*
   16: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1995.DAT:*
   17: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1996.DAT:*
   18: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1997.DAT:*
   19: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1998.DAT:*
   20: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1999.DAT:*
   21: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA2000.DAT:*
   22: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA2001A.DAT:*
   23: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA2001B.DAT:*
   24: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA2002.DAT:*

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
	1	1001	100.0	11101	11101	24	ABN83947	Human transporter
	2	731.6	73.1	2291	2291	22	AAH16506	Human cDNA sequenc
c	3	326.4	32.6	401	401	22	AAH12424	Human cDNA clone (
	4	253.4	25.3	1301	1301	21	AAH51627	Human g35018 cDNA
	5	253.4	25.3	1386	1386	21	AAH51631	Human g35018 cDNA
	6	251.1	25.1	5881	5881	22	AAL07230	Human reproductive
	7	250.8	25.1	5881	5881	23	ABL98778	Human testicular a
c	8	250.2	25.0	40433	40433	24	ABN96830	Gene #3328 used to
c	9	249.6	24.9	227968	227968	24	ABK83497	Human cDNA differe

ALIGNMENTS

	No.	Score	% Query Match	Length	DB	ID	Description
	10	249.4	24.9	5041	22	ABA17153	Human nervous syst
c	11	248.2	24.8	30110	22	AAK89230	Human digestive sy
	12	248	24.8	4374	22	AAS21306	Human cDNA sequenc
	13	247.8	24.8	2981	22	ABA17151	Human nervous syst
	14	247.8	24.8	3172	22	AAL05811	Human reproductive
	15	247.8	24.8	3172	22	AAK85276	Human immune/haema
	16	247.8	24.8	4766	23	ABL98375	Human testicular a
	17	247.8	24.8	3172	23	AAS27783	DNA encoding novel
	18	247.8	24.8	319608	22	AAS09301	Human schizophreni
	19	247.4	24.7	319608	22	AAH51601	Human chromosome 1
	20	247.2	24.7	6052	22	AAK84334	Human immune/haema
c	21	247.2	24.7	110608	24	AAK83572	Human cDNA differe
c	22	246.8	24.7	97835	24	ABK84796	Human cDNA differe
c	23	246.6	24.7	1036	21	AAC79687	Human secreted pro
	24	246.6	24.6	3249	22	ABA15833	Human nervous syst
	25	246.6	24.6	3249	23	ABL97407	Human reproductive
	26	246.6	24.6	36305	24	ABK22783	Human testicular a
c	27	246.6	24.6	5562	22	ABL03071	Human reproductive
	28	246.6	24.6	36305	24	ABA82625	Human immune/haema
	29	246.4	24.6	56933	22	ABA82625	Human high bone ma
c	30	246.4	24.6	149480	24	ABL61947	Human HBM gene reg
	31	246.4	24.6	149480	24	ABL61948	Colon adenocarcino
	32	246.4	24.6	149480	24	ABL68365	Colon adenocarcino
c	33	246.2	24.6	107820	24	AAD16230	Kidney cancer rela
	34	246.2	24.6	326791	22	AAS43104	Human ATP-binding
c	35	245.8	24.6	14525	22	AAK70511	Human Oestrogen re
	36	245.6	24.6	9589	24	ABA99282	Human ALDH5 DNA.
	37	245.6	24.5	9589	24	ABA99369	Human ALDH5 polymo
	38	245.6	24.5	21313	22	AAK82710	Human ALDH5 polymo
	39	245.6	24.5	21313	22	AAK82338	Human immune/haema
c	40	245.6	24.5	48908	22	ABK84349	Human immune/haema
c	41	245.6	24.5	222930	24	ABA15942	Human immune/haema
c	42	245.4	24.5	21989	22	ABA15942	Human immune/haema
c	43	245.4	24.5	201143	24	ABK83568	Human DNA differen
c	44	245.2	24.5	6892	22	AAL37405	Human musculoskele
c	45	245.2	24.5	32169	22	ABA14358	Human nervous syst
c		245	24.5	56743	22	AAK68202	Human immune/haema

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

RESULT 1
ABN83947

```
ID   ABN83947 standard; DNA; 11101 BP.
XX
AC   ABN83947;
XX
DT   02-OCT-2002 (first entry)
XX
DE   Human transporter protein encoding genomic DNA.
XX
KW   Human; transporter protein; zinc transporter; pharmacogenomic analysis;
KW   diagnosis; drug screening; gene therapy; kidney; testis; heart; placenta;
KW   small intestine; liver; chromosome 1; single nucleotide polymorphism;
KW   SNP; gene; ds.
XX
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FH
FT   CDS             2284..8840
FT                   /*tag= a
FT   exon            2284..2333
FT                   /product= "transporter protein"
FT                   /note= "contains 7 introns"
FT                   /*tag= b
FT                   /number= 1
FT   intron          2334..2962
FT                   /*tag= c
FT                   /number= 1
FT   exon            2963..3183
FT                   /*tag= d
FT                   /number= 2
FT   intron          3184..3735
```

```
FT   exon
FT             /*tag= e
FT             3736..3882
FT             /*tag= f
FT             /number= 3
FT   intron
FT             3883..4629
FT             /*tag= g
FT             /number= 3
FT   exon
FT             4630..4783
FT             /*tag= h
FT             /number= 4
FT   intron
FT             4784..5472
FT             /*tag= i
FT             /number= 4
FT   exon
FT             5473..5632
FT             /*tag= j
FT             /number= 5
FT   intron
FT             5633..6375
FT             /*tag= k
FT             /number= 5
FT   exon
FT             6376..6481
FT             /*tag= l
FT             /number= 6
FT   intron
FT             6482..8088
FT             /*tag= m
FT             /number= 6
FT   exon
FT             8089..8223
FT             /*tag= n
FT             /number= 7
FT   intron
FT             8224..8694
FT             /*tag= o
FT             /number= 7
FT   exon
FT             8695..8838
FT             /*tag= p
FT             /number= 8
XX
PN   WO20224910-A2.
XX
PD   28-MAR-2002.
XX
PF   20-SEP-2001; 2001WO-US29218.
XX
PR   20-SEP-2000; 2000US-234160P.
PR   19-OCT-2000; 2000US-691219.
XX
PA   (PEKE ) PE CORP NY.
XX
PI   Wei M, Ketchum KA, Di Francesco V, Beasley EM;
XX
DR   WPI; 2002-404954/43.
DR   P-PSDB; ABB83084.
XX
PT   Novel human transporter proteins, related to zinc transporter
PT   subfamily, useful as model for developing human therapeutic targets and
PT   serves as target for human therapeutics.
XX
PS   Claim 4(b); Fig 3; 75pp; English.
XX
CC   The invention relates to an isolated human transporter protein that is
CC   related to the zinc transporter subfamily. Polynucleotides and
CC   polypeptides of the invention are useful for treating a disease or
CC   condition mediated by human transporter protein. The proteins also
CC   provide a target for diagnosing a disease or predisposition to disease
CC   mediated by the peptide, and in pharmacogenomic analysis. The peptides
CC   are also useful for treating a disorders characterised by absence of,
CC   inappropriate or unwanted expression of the protein. The nucleic acids
CC   are also useful in drug screening assays and as a target for treatment
CC   by the compounds identified through drug screening. The invention also
CC   provides vectors for gene therapy in patients with aberrant expression
CC   of the gene encoding the transporter protein. The gene of the invention
CC   has been found to be expressed in humans in the kidney, testis, heart,
CC   placenta, small intestine and liver. The gene has been localised to
CC   human chromosome 1. The current sequence represents the human transporter
```

```
CC   protein encoding genomic DNA.
CC   Note: This sequence contains 6 single nucleotide polymorphisms (SNP's),
CC   but information given in the specification is insufficient to determine
CC   their locations within this sequence (see ABN83948-ABN83953 for specific
CC   regions).
XX
SQ   Sequence 11101 BP; 2403 A; 3041 C; 3200 G; 2448 T; 9 other;

Query Match             100.0%;  Score 1001;  DB 24;  Length 11101;
Best Local Similarity   100.0%;  Pred. No. 5.4e-237;
Matches 1001;  Conservative  0;  Mismatches  0;  Indels  0;  Gaps  0;
```

```
Qy        1 TTACTTTTACCATCCTGGGGACAGGAGGAACTCTGGTCCGGAGGCTGCCTCACTGAGGAGT 60
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    10000 TTACTTTTACCATCCTGGGGACAGGAGGAACTCTGGTCCGGAGGCTGCCTCACTGAGGAGT 10059

Qy       61 CAGGTGGGCGCTTCCCAGGGCAGTGCCAGTGCTGCTGAGTTTGTTGTTGACTGTTCTGGCCC 120
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    10060 CAGGTGGGCGCTTCCCAGCCTCCCAGGGCCAGTGCTGCTGAGTTTGTTGTTGACTGTTCTGGCCC 10119

Qy      121 AAGGTGGGGAGGAGGTGGGGTTGGGTGTCACTTGGCCTCCCCACTTTAAATCTCTGTCTTTCCATC 180
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    10120 AAGGTGGGGAGGAGGTGGGGTTGGGTGTCACTTGGCCTCCCCACTTTAAATCTCTGTCTTTCCATC 10179

Qy      181 TGTGAAATGACCTCTTTTGTGCCTTCCCCAGCCACTGTCATCCTGAAAAGTCACTGGCCTGGT 240
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    10180 TGTGAAATGACCTCTTTTGTGCCTTCCCCAGCCACTGTCATCCTGAAAAGTCACTGGCCTGGT 10239

Qy      241 AGGTGGGGTGCCTTCAGCCCCTCCAGGTCGTGTGAAAAGTCTGTGAAAGCACTGGCCTGGAG 300
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    10240 AGGTGGGGTGCCTTCAGCCCCTCCAGGTCGTGTGAAAAGTCTGTGAAAGCACTGGCCTGGAG 10299

Qy      301 AGGGGTGGGGGGGTTGGCTGGGGTGTGGGCTCCATTCCACACAAATCTCAGGGGACTCAACCTC 360
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    10300 AGGGGTGGGGGGGTTGGCTGGGGTGTGGGCTCCATTCCACACAAATCTCAGGGGACTCAACCTC 10359

Qy      361 CCCTACCCAAACTACCCCACCCCACCCCAAGCCAATGGCAGGCCCCCAGGGAACTTGATCCTGG 420
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    10360 CCCTACCCAAACTACCCCACCCCACCCCAAGCCAATGGCAGGCCCCCAGGGAACTTGATCCTGG 10419

Qy      421 GCTTTGCCGTATGCCAAGTCCTTTACACCCCTCTCAAGAGCAGCAGTCAATTGGCCTGGGCACGG 480
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    10420 GCTTTGCCGTATGCCAAGTCCTTTACACCCCTCTCAAGAGCAGCAGTCAATTGGCCTGGGCACGG 10479

Qy      481 TGGCTCATGCCTGCAATCCCAGCACCTTGGGAGGCCGAGGCAGATGACTTGAGGCC 540
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    10480 TGGCTCATGCCTGCAATCCCAGCACCTTGGGAGGCCGAGGCAGATGACTTGAGGCC 10539

Qy      541 AGGAGTTCGAGACCAGCCTGGGCCAATATGGCGGAAAACCTCATTTCTACTAAAAATTAAACAAA 600
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    10540 AGGAGTTCGAGACCAGCCTGGGCCAATATGGCGGAAAACCTCATTTCTACTAAAAATTAAACAAA 10599

Qy      601 ACTAACCAGGCGTGGTGGCTGGCGCCTGTAATCCCAGCTACTCGGGAGGCTGAGGCAGGA 660
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    10600 ACTAACCAGGCGTGGTGGCTGGCGCCTGTAATCCCAGCTACTCGGGAGGCTGAGGCAGGA 10659

Qy      661 GAATCGCTTGAACCGGGAGGCGGAGGTTGCAGTGAGCTGAGATCACCACCACTGCACTCC 720
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    10660 GAATCGCTTGAACCGGGAGGCGGAGGTTGCAGTGAGCTGAGATCACCACCACTGCACTCC 10719

Qy      721 AGCCTGGGCGACAGAGCGAGACTCCAGCTTAAAAAAAAAAAAAAAAAAAAAAGGAGA 780
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    10720 AGCCTGGGCGACAGAGCGAGACTCCAGCTTAAAAAAAAAAAAAAAAAAAAAAGGAGA 10779

Qy      781 CCATCACTGCCGTCCTGCATTCTTACAGATGAAAAAACAGGGCTTCAGAGGTTGAATCGTTT 840
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    10780 CCATCACTGCCGTCCTGCATTCTTACAGATGAAAAAACAGGGCTTCAGAGGTTGAATCGTTT 10839

Qy      841 TCCTGAAGTCAGACAGCCAGTGCAGGCAGGTCTGGGGATTTCTGCCTCATTTCGGTAGACC 900
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    10840 TCCTGAAGTCAGACAGCCAGTGCAGGCAGGTCTGGGGATTTCTGCCTCATTTCGGTAGACC 10899

Qy      901 TTCCTCTACAGCAGAGGGTCTGGGGGGCCCTGTGGGTCTGCCGTGTTGGTACAATACAAA 960
```

```
Db   10900 TTCCTCTACAGCAGGGTCGCCGCGCCGCCCATGGGGTGAGGACATG 1001
           ||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy     961 CCCCTGGGAGCCAGCCAGTGCGCGGCCGGGCCTGGTCGGGTGCGCGGCCGGGCCTGGTCGGGTGCGCGGCCGGGCCTGGTCGGGTGCTGGTCGGGTGTTGGTACAATACAAA 10959
           ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   10960 CCCCTGGGAGCCAGCCAGTGCCGGGCCCCATGGGTGAGGACATG 11000
```

RESULT 2

AAH16506
ID   AAH16506 standard; cDNA; 2291 BP.
XX
AC   AAH16506;
XX
DT   26-JUN-2001 (first entry)
XX
DE   Human cDNA sequence SEQ ID NO:15542.
XX
KW   Human; primer; detection; diagnosis; antisense therapy; gene therapy; ss.
XX
OS   Homo sapiens.
XX
PN   EP1074617-A2.
XX
PD   07-FEB-2001.
XX
PF   28-JUL-2000; 2000EP-0116126.
XX
PR   29-JUL-1999; 99JP-0248036.
PR   27-AUG-1999; 99JP-0300253.
PR   11-JAN-2000; 2000JP-0018776.
PR   02-MAY-2000; 2000JP-0183767.
PR   09-JUN-2000; 2000JP-0241899.
XX
PA   (HELI-) HELIX RES INST.
XX
PI   Ota T, Isogai T, Nishikawa T, Hayashi K, Saito K, Yamamoto J;
PI   Ishii S, Sugiyama T, Wakamatsu A, Nagai K, Otsuki T;
XX
DR   WPI; 2001-318749/34.
XX
PT   Primer sets for synthesizing polynucleotides, particularly the 5602
PT   full-length cDNAs defined in the specification, and for the detection
PT   and/or diagnosis of the abnormality of the proteins encoded by the
PT   full-length cDNAs -
XX
PS   Claim 8; SEQ ID 15542; 2537pp + CD ROM; English.
XX
CC   The present invention describes primer sets for synthesising 5602
CC   full-length cDNAs defined in the specification. Where a primer set
CC   comprises: (a) an oligo-dT primer and an oligonucleotide complementary
CC   to the complementary strand of a polynucleotide which comprises one of
CC   the 5602 nucleotide sequences defined in the specification, where the
CC   oligonucleotide comprises at least 15 nucleotides; or (b) a combination
CC   of an oligonucleotide comprising a sequence complementary to the
CC   complementary strand of a polynucleotide which comprises a 5'-end
CC   sequence and an oligonucleotide comprising a sequence complementary to a
CC   polynucleotide which comprises a 3'-end sequence, where the
CC   oligonucleotide comprises at least 15 nucleotides and the combination of
CC   the 5'-end sequence/3'-end sequence is selected from those defined in
CC   the specification. The primer sets can be used in antisense therapy and
CC   in gene therapy. The primers are useful for synthesizing polynucleotides,
CC   particularly full-length cDNAs. The primers are also useful for the
CC   detection and/or diagnosis of the abnormality of the proteins encoded by
CC   the full-length cDNAs. The primers allow obtaining of the full-length
CC   cDNAs easily without any specialised methods. AAH03166 to AAH13628 and
CC   AAH13633 to AAH18742 represent human cDNA sequences; AAB92446 to
CC   AAH13629 to AAH13632
CC   AAB95893 represent human amino acid sequences; and AAH13629 to AAH13632
CC   represent oligonucleotides, all of which are used in the exemplification
CC   of the present invention.
XX
SQ   Sequence 2291 BP; 474 A; 663 C; 647 G; 507 T; 0 other;
```

Query Match   73.1%; Score 731.6; DB 22; Length 2291;
Best Local Similarity   99.3%; Pred. No. 9.9e-171;
Matches 745; Conservative 0; Mismatches 4; Indels 1; Gaps 1;

```
Qy      1 TTACTTTTACCATTCCTGGGGACAGCGGGAAACTCTGTGCCCGAGGCTGCCTCACTGAGGAGT 60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1543 TTACTTTTACCATTCCTGGGGACAGCGGGAAACTCTGTGCCCGAGGCTGCCTCACTGAGGAGT 1602

Qy     61 CAGGTGGGCTTCCCAGCCTCCCCAGGGCACGCACTGCTGAGTTTGTTCTTGACTGTTCTCGGCCC 120
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1603 CAGGTGGGCTTCCCAGCCTCCCCAGGGCACGCACTGCTGAGTTTGTTCTTGACTGTTCTCGGCCC 1662

Qy    121 AAGGTGGGAGGAGGAGGTGGGGGTTTGGTCACTTGCCTCCCCACTTTAAAATCTCGTGTCTTCCATC 180
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1663 GAGGTGGGAGGAGGAGGTGGGGGTTTGGTCACTTGCCTCCCCACTTTAAAATCTCGTGTCTTCCATC 1722

Qy    181 TGTGAAATGACCTCTTTTGTGCCTTCCCAGCACTGTCCCCACACTGTGTGTTCTTAGGT 240
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1723 TGTGAAATGACCTCTTTTGTGCCTTCCCAGCACTGTCCCCACACTGTGTGTTCTTAGGT 1782

Qy    241 AGGTGGGGTCTTCAGCCCCCTCCAGGTCTGTGAAAAGTCTGTGGGAAAGCACTGGGCCTGGGAG 300
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1783 AGATGGGGTCTTCAGCCCCCTCCAGGTCTGTGAAAAGTCTGTGGGAAAGCACTGGGCCTGGGAG 1842

Qy    301 AGGGGGTGGGGGGGTTGCTGGTGGGGTGGGGTGGGTGCTCCATTCCACCACCAAATCTCAGGGGGACTCAACCTC 360
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1843 AGGGGGTGGGGGGGTTGCTGGTGGGGTGGGGTGGGTGCTCCATTCCACCACCAAATCTCAGGGGGACTCAACCTC 1902

Qy    361 CCCTACCCAACTACCCCACCCCCCACCCCAACCCCATGGCCAGGCCCCAGGAACTTGATCCTGG 420
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1903 CCCTACCCAACT--CCCCACCCCCCACCCCAACCCCATGGCCAGGCCCCAGGAACTTGATCCTGG 1961

Qy    421 GCTTTGCCGTATGCCAAGTCCTTACACCCCCTCTCAAGAGACACAGTCATTGCTGGGCGACGG 480
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1962 GCTTTGCCGTATGCCAAGTCCTTACACCCCCTCTCAAGAGACACAGTCATTGCTGGGCGACGG 2021

Qy    481 TGGCTCATGCCTGCAATCCCAGCACACCCTTGGGAGGCTGAGGCAGGAGGATGACTTGAGGCC 540
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2022 TGGCTCATGCCTGCAATCCCAGCACACCCTTGGGAGGCTGAGGCAGGAGGATGACTTGAGGCC 2081

Qy    541 AGGAGTTCGAGACCAGCCTGGCCAATATGGGGGAAAACCTCATTTCTACTAAAAATACAAAA 600
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2082 AGGAGTTCGAGACCAGCCTGGCCAATATGGGGGAAAACCTCATTTCTACTAAAAATACAAAA 2141

Qy    601 ACTAACCAGGCGCGTGGTGGCGCGTGCCTGTGTAATCCCAGCTACTCACTCGGGAGGCTGAGGCAGGA 660
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2142 ACTAACCAGGCGCGTGGTGGCGCGTGCCTGTGTAATCCCAGCTACTCGGGAGGCTGAGGCAGGA 2201

Qy    661 GAATCGCTTGAACCGGGGAGGCAGAGGTTGCAGTGAGCTGAGATCACGCCACTGCACTCC 720
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2202 GAATCGCTTGAACCGGGGAGGCAGAGGTTGCAGTGAGCTGAGATCACACCACTGCACTCC 2261

Qy    721 AGCCTGGGCGACAGAGCGAGACTCCAGCTT 750
          ||||||||||||||||||||||||||||||
Db   2262 AGCCTGGGCGACAGAGCAAGACTCCAGCTT 2291
```

RESULT 3

AAH12422/c
ID   AAH12422 standard; cDNA; 401 BP.
XX
AC   AAH12422;
XX
DT   26-JUN-2001 (first entry)
XX
DE   Human cDNA clone (3'-primer) SEQ ID NO:9257.
XX
KW   Human; primer; detection; diagnosis; antisense therapy; gene therapy; ss.
XX
OS   Homo sapiens.
XX
PN   EP1074617-A2.
XX

```
PD   07-FEB-2001.
XX
PF   28-JUL-2000; 2000EP-0116126.
XX
PR   29-JUL-1999; 99JP-0248036.
PR   27-AUG-1999; 99JP-0300253.
PR   11-JAN-2000; 2000JP-0118776.
PR   02-MAY-2000; 2000JP-083767.
PR   09-JUN-2000; 2000JP-0241899.
XX
XX
PA   (HELI-) HELIX RES INST.
XX
PI   Ota T, Isogai T, Nishikawa T, Hayashi K, Saito K, Yamamoto J;
PI   Ishii S, Sugiyama T, Wakamatsu A, Nagai K, Otsuki T;
XX
DR   WPI; 2001-318749/34.
XX
PT   Primer sets for synthesizing polynucleotides, particularly the 5602
PT   full-length cDNAs defined in the specification, and for the detection
PT   and/or diagnosis of the abnormality of the proteins encoded by the
PT   full-length cDNAs -
XX
PS   Claim 3; SEQ ID 9257; 2537pp + CD ROM; English.
XX
CC   The present invention describes primer sets for synthesizing 5602
CC   full-length cDNAs defined in the specification. Where a primer set
CC   comprises: (a) an oligo-dT primer and an oligonucleotide complementary
CC   to the complementary strand of a polynucleotide which comprises one of
CC   the 5602 nucleotide sequences defined in the specification, where the
CC   oligonucleotide comprises at least 15 nucleotides; or (b) a combination
CC   of an oligonucleotide comprising a sequence complementary to the
CC   complementary strand of a polynucleotide which comprises a 5'-end
CC   sequence and an oligonucleotide comprising a sequence complementary to a
CC   polynucleotide which comprises a 3'-end sequence, where the
CC   oligonucleotide comprises at least 15 nucleotides and the combination of
CC   the 5'-end sequence/3'-end sequence is selected from those defined in
CC   the specification. The primer sets can be used in antisense therapy and
CC   in gene therapy. The primers are useful for synthesizing polynucleotides,
CC   particularly full-length cDNAs. The primers are also useful for the
CC   detection and/or diagnosis of the abnormality of the proteins encoded by
CC   the full-length cDNAs. The primers allow obtaining of the full-length
CC   cDNAs easily without any specialised methods. AAH03166 to AAH13628 and
CC   AAH13633 to AAH18742 represent human cDNA sequences; AAB52446 to
CC   AAB95893 represent human amino acid sequences; and AAH13629 to AAH13632
CC   represent oligonucleotides, all of which are used in the exemplification
CC   of the present invention.
XX
SQ   Sequence 401 BP; 87 A; 109 C; 100 G; 96 T; 9 other;

Query Match             32.6%;  Score 326.4;  DB 22;  Length 401;
Best Local Similarity   93.2%;  Pred. No. 5.5e-71;
Matches   358;  Conservative   0;  Mismatches   24;  Indels    2;  Gaps    2;

Qy   368 CAACTACCCCACCCCACCCCAAGCCATGGCAGGCCCCAGGGAACTTGAATCCTCGGGCTTTGC 427
         || ||| |||| || |||||||||||||||||||||||||||||| |||||||| |||||||
Db   383 CANTTTCCAACCCNACCCCAAGCCATGGCCAGGCCCCAGGGAAATTGAATCCTTGGTTTTTGC 324

Qy   428 CGTATGGCCAAGTCNTTACACCCTCTCAAGAGACAACAGTCATTGGCTCGGGCACGGTGGCTCA 487
         |||||||||||||||| |||||||||||||||||||||||||||||||||||||||||| |||
Db   323 NGTATCCCAAGTCNTTACACCCTTNTCAAGAGACAACAGTCATTGGCTCGGGCACGGTGGTTCA 264

Qy   488 TGCCTGCGAATCCCAGCACCTTGGGAGGCTGAGGCAGGCAGATCAGTTGAGGCCCAGGAGTT 547
         |||||| ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   263 TGCNTGCAATCCCAGCCACCTTGGGAGGCTGAGGCAGGCGAGATGACTTGAGGCCCAGGAGTT 204

Qy   548 CGAGACCAGCCTGGGCCAANTATGGCGAAACCTCATTGTGTCTAAAAAATTACAAAAACTAACC 607
         |||||||||||||||| ||||||||||||||||||||||||||||||||||||||||||||
Db   203 CGAGACCAGCCTGGCCAATATGGCGAAACNTCATTTTTACTAAAAATACAAAAAACTAACC 144

Qy   608 AGGCGTGGTGGCGTGGGTCGTGCCCTGTGAAT-CCCAGCTACTCGGGAGGCTGAGGCAGGAGAATCG 666
         ||||||||||||||||||| ||||||||||||||| |||||||||||||||||||||||||||||
Db   143 AGGCGTGGTGGCTGGTGCCTGTGTCCTGTAATCCCAGCTACTCGGGAGGCTGAGGCAGGAGNTTCG 84
```

```
RESULT 4
AAH51627
ID   AAH51627 standard; cDNA; 1301 BP.
XX
AC   AAH51627;
XX
DT   29-AUG-2001  (first entry)
XX
DE   Human g35018 cDNA sequence SEQ ID 36.
XX
KW   sbg1; g34665; sbg2; g35017; g35018; chromosome 13q31-q33; haplotype;
KW   biallelic marker; polymorphism; schizophrenia; bipolar disorder; ss.
XX
OS   Homo sapiens.
XX
XX
PN   WO200058510-A2.
XX
PD   05-OCT-2000.
XX
PF   30-MAR-2000; 2000WO-IB00435.
XX
PR   30-MAR-1999; 99US-0126903.
PR   30-APR-1999; 99US-0131971.
PR   30-APR-1999; 99US-0132065.
PR   14-JUL-1999; 99US-0143928.
PR   27-JUL-1999; 99US-0145915.
PR   29-JUL-1999; 99US-0146452.
PR   29-JUL-1999; 99US-0146453.
PR   28-OCT-1999; 99US-0162288.
XX
PA   (GEST ) GENSET.
XX
PI   Cohen D, Blumenfeld M, Chumakov I, Bougueleret L, Bihain B;
PI   Essioux L;
XX
DR   WPI; 2000-619082/59.
XX
PT   Polynucleotides comprising sequences from sbg1 and g35018 biallelic
PT   markers are used for genotyping and detecting schizophrenia or bipolar
PT   disorder and predisposition to these disorders -
XX
PS   Claim 2; Page 550-551; 737pp; English.
XX
CC   AAH51601 represents a human genomic nucleotide sequence comprising sbg1,
CC   g34665, sbg2, g35017 and g35018 nucleic acid sequences located on the
CC   human chromosome 13q31-q33 locus. The nucleotide sequences contain
CC   biallelic markers and polymorphisms. Sequences AAH51602 - AAH51626 and
CC   AAB62907 - AAB62915 represent cDNA human sbg1 cDNA sequences and protein
CC   products. AAH51627 - AAH51631 and AAB62916 - AAB62918 represent g35018
CC   cDNA sequences and protein products. Primers AAH51632 - AAH51699 are used
CC   to isolate sbg1 cDNAs, while sbg1 exons AAH51633 - AAH51699 are
CC   represented by sequences AAH51642 - AAH51699. Nucleotide sequences of
CC   amplicons which comprise biallelic markers AAH51700 - AAH51817. Biallelic markers
CC   13q31-q33 locus are represented in the sequences by degenerate/undefined base codes. PCR
CC   primers AAH51818 and AAH51819 are used in the isolation of sequences of
CC   the invention. The biallelic marker containing nucleotide sequences are
CC   used to determine the identity of the nucleotide at a biallelic marker in
CC   a sample DNA sequence. The nucleotide sequences may be labelled and used
CC   for genotyping by determining the identity of a nucleotide at a Region
CC   D-related biallelic marker in a biological sample from single or multiple
CC   subjects. By determining the frequency of a biallelic marker in a
CC   population an association between a genotype and a trait, a haplotype and
CC   a trait and a phenotype and a trait can be detected. The sequences can be
```

```
Qy   667 CTTGAACCGGGAGGCGAGGAGGGTTGCAGTGAGGAGATCACCACGTGCCACTCCAGCCTG 726
         |||||||| ||||||||| |||||||||||||||| |||||||||||||| |||||||| ||
Db    83 CTTGAACCGGGAGGCGAGGCAGAGGGTTGCAGTGAGCGTGAGATCACCACCATTGCCAGCTTG 24

Qy   727 GGGCACAGAGCGCAGACTCCAGCTT 750
         || |||||| || |||||||||
Db    23 GG-GACAGAGCAGATTCCAGCTT 1
```

CC   used to determine a predisposition to or early onset of schizophrenia or
CC   bipolar disorder or a beneficial response to or side effects related to
CC   treatment against schizophrenia or bipolar disorder.
XX
SQ   Sequence 1301 BP; 411 A; 292 C; 276 G; 310 T; 12 other;

Query Match   Score 253.4;   DB 21;   Length 1301;
Best Local Similarity   84.3%;   Pred. No. 7.8e-53;
Matches 284;   Conservative   1;   Mismatches 52;   Indels 0;   Gaps 0;

Qy   444 ACACCCCTCTCAAGAGACAGTCATTTGGCTCGGGCACCGTGGGCTTCATCCCTGCAATCCCAGC 503
          ||||| ||||||| |||||||||||||||||||||| |||| ||||| | |||||||||||||||
Db   492 ACCCTGCTTTTAAGAGAGGAAGGTCTTTGGCCAGGGCTCAGTGGTCTCACACCCTGTAATCCCAGC 551

Qy   504 ACCTTGGGAGGCTGAGGCAGGCAGATGACTTGAGCCAGGAGTTCGAGACCAGCCTGGCC 563
          ||||||||||||||||||||||| |||||||||| ||||||||||| |||||||||||||
Db   552 ACTTTGGGAGGCCGAGGCGGGCAGATCACCTGAGGTCAGGAGTTCAAGACCAGCCTGACC 611

Qy   564 AATATGGCGAAACCTCATTTTCTACTAAAAATACAAAAATTAACCACCAGCGTGGCTGGCTTGT 623
          |||||| ||||||||||||| ||||||||||||||||||||||| |||||||||||||||||||
Db   612 AATATGGTGAAACCCCATCTCTACTAAAAATTACAAAAATTAGCCAGGCATGGTGGCACGGT 671

Qy   624 GCCTGTAATCCCAGCTACTCGGGAGGCTGAGGCAGGAGAATCGCTTGAACCGGGGAGGCA 683
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   672 GCCTGTAGTCCCAGCTACTCGGGAGGCTGAGACAGGAGAATTGCTTGAACCTGGGAGGCG 731

Qy   684 GAGGTTGCAGTGAGCTGAGATCACACCACTGCACTCCAGCCTGGGCGACAGAGCGAGACT 743
          ||||||||||||||||||||||||||||||||||||||||||||||||| ||||||||||
Db   732 GAGGTTGCAGTGAGCCAAGATCACACCACTGCACTCCAGCCTGGGCAACAGAGTGAGACT 791

Qy   744 CCAGCCTTAAAAAAAAAAAAAAAAAAAAAAAAAAAAAGAGA 780
          || ||| ||:|:||||||||||||||||||||| ||| | |
Db   792 CCATCTCAAATTAARAAAAAAAATAAGATACAGAATA 828

RESULT 5
AAH51631
ID   AAH51631 standard; cDNA; 1386 BP.
XX
AC   AAH51631;
XX
DT   29-AUG-2001 (first entry)
XX
DE   Human g35018 cDNA sequence SEQ ID 40.
XX
KW   sbg1; 934665; sbg2; 935017; 935018; chromosome 13q31-q33; haplotype;
KW   biallelic marker; polymorphism; schizophrenia; bipolar disorder; ss.
XX
OS   Homo sapiens.
XX
PN   WO200058510-A2.
XX
PD   05-OCT-2000.
XX
PF   30-MAR-2000; 2000WO-IB00435.
XX
PI   Cohen D, Blumenfeld M, Chumakov I, Bougueleret L, Bihain B;
PI   Essioux L;
PA   (GENST) GENSET.
XX
PR   30-MAR-1999;   99US-0126903.
PR   30-APR-1999;   99US-0131971.
PR   30-APR-1999;   99US-0132065.
PR   14-JUL-1999;   99US-0143928.
PR   27-JUL-1999;   99US-0145915.
PR   29-JUL-1999;   99US-0146452.
PR   29-JUL-1999;   99US-0146453.
PR   28-OCT-1999;   99US-0162288.
XX
DR   WPI; 2000-619082/59.
XX
PT   Polynucleotides comprising sequences from sbg1 and g35018 biallelic

PT   markers are used for genotyping and detecting schizophrenia or bipolar
PT   disorder and predisposition to these disorders -
XX
PS   Claim 2; Page 558; 737pp; English.
XX
CC   AAH51601 represents a human genomic nucleotide sequence comprising sbg1,
CC   934665, sbg2, 935017 and 935018 nucleic acid sequences located on the
CC   human chromosome 13q31-q33 locus. The nucleotide sequences contain
CC   biallelic markers and polymorphism. Sequences AAH51642 - AAH51699 and
CC   AAB62907 - AAB62915 represent cDNA human sbg1 cDNA sequences and protein
CC   products. AAH51627 - AAH51631 and AAB62916 - AAB62918 represent g35018
CC   cDNA sequences and protein products. Primers AAH51632 - AAH51699 are used
CC   to isolate sbg1 cDNAs, while sbg1 exons from different primers are
CC   represented by sequences AAH51642 - AAH51699. Nucleotide sequences of
CC   amplicons which comprise biallelic markers located on the chromosome
CC   13q31-q33 locus are represented in the sequences by degenerate/undefined base codes. PCR
CC   primers AAH51818 and AAH51819 are used in the isolation of sequences of
CC   the invention. The biallelic marker containing nucleotide sequences are
CC   used to determine the identity of the nucleotide at a biallelic marker in
CC   a sample DNA sequence. The nucleotide sequences may be labelled and used
CC   for genotyping by determining the identity of a nucleotide at a Region
CC   D-related biallelic marker in a biological sample from single or multiple
CC   subjects. By determining the frequency of a biallelic marker in a
CC   population an association between a genotype and a trait, a haplotype and
CC   a trait and a phenotype and a trait can be detected. The sequences can be
CC   used to determine a predisposition to or early onset of schizophrenia or
CC   bipolar disorder or a beneficial response to or side effects related to
CC   treatment against schizophrenia or bipolar disorder.
XX
SQ   Sequence 1386 BP; 319 C; 297 G; 332 T; 12 other;

Query Match   Score 253.4;   DB 21;   Length 1386;
Best Local Similarity   84.3%;   Pred. No. 8e-53;
Matches 284;   Conservative   1;   Mismatches 52;   Indels 0;   Gaps 0;

Qy   444 ACACCCCTCTCAAGAGACAGTCATTTGGCTCGGGCACCGTGGGCTTCATCCCTGCAATCCCAGC 503
          ||||| ||||||| |||||||||||||||||||||| |||| ||||| | |||||||||||||||
Db   577 ACCCTGCTTTTAAGAGAGGAAGGTCTTTGGCCAGGGCTCAGTGGTCTCACACCCTGTAATCCCAGC 636

Qy   504 ACCTTGGGAGGCTGAGGCAGGCAGATGACTTGAGCCAGGAGTTCGAGACCAGCCTGGCC 563
          ||||||||||||||||||||||| |||||||||| ||||||||||| |||||||||||||
Db   637 ACTTTGGGAGGCCGAGGCGGGCAGATCACCTGAGGTCAGGAGTTCAAGACCAGCCTGACC 696

Qy   564 AATATGGCGAAACCTCATTTTCTACTAAAAATACAAAAATTAACCACCAGCGTGGCTGGCTTGT 623
          |||||| ||||||||||||| ||||||||||||||||||||||| |||||||||||||||||||
Db   697 AATATGGTGAAACCCCATCTCTACTAAAAATTACAAAAATTAGCCAGGCATGGTGGCACGGT 756

Qy   624 GCCTGTAATCCCAGCTACTCGGGAGGCTGAGGCAGGAGAATCGCTTGAACCGGGGAGGCA 683
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   757 GCCTGTAGTCCCAGCTACTCGGGAGGCTGAGACAGGAGAATTGCTTGAACCTGGGAGGCG 816

Qy   684 GAGGTTGCAGTGAGCTGAGATCACACCACTGCACTCCAGCCTGGGCGACAGAGCGAGACT 743
          ||||||||||||||||||||||||||||||||||||||||||||||||| ||||||||||
Db   817 GAGGTTGCAGTGAGCCAAGATCACACCACTGCACTCCAGCCTGGGCAACAGAGTGAGACT 876

Qy   744 CCAGCCTTAAAAAAAAAAAAAAAAAAAAAAAAAAAAAGAGA 780
          || ||| ||:|:||||||||||||||||||||| ||| | |
Db   877 CCATCTCAAATTAARAAAAAAAATAAGATACAGAATA 913

RESULT 6
AAL07230
ID   AAL07230 standard; DNA; 5881 BP.
XX
AC   AAL07230;
XX
DT   21-NOV-2001 (first entry)
XX
DE   Human reproductive system related antigen DNA SEQ ID NO: 9918.
XX
KW   Human; reproductive system related antigen; reproductive system disorder;
KW   cancer; gene therapy; ds.

```
XX
OS   Homo sapiens.
XX
PN   WO200155320-A2.
XX
PD   02-AUG-2001.
XX
PF   17-JAN-2001;   2001WO-US01339.
XX
PR   31-JAN-2000;   2000US-0179065.
PR   04-FEB-2000;   2000US-0180628.
PR   24-FEB-2000;   2000US-0184664.
PR   02-MAR-2000;   2000US-0186350.
PR   16-MAR-2000;   2000US-0189874.
PR   17-MAR-2000;   2000US-0190076.
PR   18-APR-2000;   2000US-0198123.
PR   19-MAY-2000;   2000US-0205515.
PR   07-JUN-2000;   2000US-0209467.
PR   28-JUN-2000;   2000US-0214886.
PR   30-JUN-2000;   2000US-0215135.
PR   07-JUL-2000;   2000US-0216647.
PR   07-JUL-2000;   2000US-0216880.
PR   11-JUL-2000;   2000US-0217487.
PR   11-JUL-2000;   2000US-0217496.
PR   14-JUL-2000;   2000US-0218290.
PR   26-JUL-2000;   2000US-0220963.
PR   26-JUL-2000;   2000US-0220964.
PR   14-AUG-2000;   2000US-0224418.
PR   14-AUG-2000;   2000US-0224519.
PR   14-AUG-2000;   2000US-0225213.
PR   14-AUG-2000;   2000US-0225214.
PR   14-AUG-2000;   2000US-0225266.
PR   14-AUG-2000;   2000US-0225267.
PR   14-AUG-2000;   2000US-0225268.
PR   14-AUG-2000;   2000US-0225270.
PR   14-AUG-2000;   2000US-0225447.
PR   14-AUG-2000;   2000US-0225757.
PR   14-AUG-2000;   2000US-0225758.
PR   14-AUG-2000;   2000US-0225759.
PR   18-AUG-2000;   2000US-0226279.
PR   22-AUG-2000;   2000US-0226681.
PR   22-AUG-2000;   2000US-0226868.
PR   22-AUG-2000;   2000US-0227009.
PR   23-AUG-2000;   2000US-0227182.
PR   30-AUG-2000;   2000US-0228924.
PR   01-SEP-2000;   2000US-0229287.
PR   01-SEP-2000;   2000US-0229343.
PR   01-SEP-2000;   2000US-0229344.
PR   01-SEP-2000;   2000US-0229345.
PR   05-SEP-2000;   2000US-0229509.
PR   05-SEP-2000;   2000US-0229513.
PR   06-SEP-2000;   2000US-0230437.
PR   06-SEP-2000;   2000US-0230438.
PR   08-SEP-2000;   2000US-0231242.
PR   08-SEP-2000;   2000US-0231243.
PR   08-SEP-2000;   2000US-0231244.
PR   08-SEP-2000;   2000US-0231413.
PR   08-SEP-2000;   2000US-0231414.
PR   08-SEP-2000;   2000US-0232080.
PR   08-SEP-2000;   2000US-0232081.
PR   12-SEP-2000;   2000US-0231968.
PR   14-SEP-2000;   2000US-0232397.
PR   14-SEP-2000;   2000US-0232398.
PR   14-SEP-2000;   2000US-0232399.
PR   14-SEP-2000;   2000US-0232400.
PR   14-SEP-2000;   2000US-0232401.
PR   14-SEP-2000;   2000US-0233063.
PR   14-SEP-2000;   2000US-0233064.
PR   14-SEP-2000;   2000US-0233065.
PR   21-SEP-2000;   2000US-0234223.
PR   21-SEP-2000;   2000US-0234274.
PR   25-SEP-2000;   2000US-0234997.
PR   25-SEP-2000;   2000US-0234998.

PR   26-SEP-2000;   2000US-0235484.
PR   27-SEP-2000;   2000US-0235834.
PR   27-SEP-2000;   2000US-0235836.
PR   29-SEP-2000;   2000US-0236327.
PR   29-SEP-2000;   2000US-0236367.
PR   29-SEP-2000;   2000US-0236368.
PR   29-SEP-2000;   2000US-0236369.
PR   29-SEP-2000;   2000US-0236370.
PR   02-OCT-2000;   2000US-0236802.
PR   02-OCT-2000;   2000US-0237037.
PR   02-OCT-2000;   2000US-0237038.
PR   02-OCT-2000;   2000US-0237039.
PR   02-OCT-2000;   2000US-0237040.
PR   13-OCT-2000;   2000US-0239935.
PR   13-OCT-2000;   2000US-0239937.
PR   20-OCT-2000;   2000US-0240960.
PR   20-OCT-2000;   2000US-0241221.
PR   20-OCT-2000;   2000US-0241785.
PR   20-OCT-2000;   2000US-0241786.
PR   20-OCT-2000;   2000US-0241787.
PR   20-OCT-2000;   2000US-0241808.
PR   20-OCT-2000;   2000US-0241809.
PR   20-OCT-2000;   2000US-0241826.
PR   01-NOV-2000;   2000US-0244617.
PR   08-NOV-2000;   2000US-0246474.
PR   08-NOV-2000;   2000US-0246475.
PR   08-NOV-2000;   2000US-0246476.
PR   08-NOV-2000;   2000US-0246477.
PR   08-NOV-2000;   2000US-0246478.
PR   08-NOV-2000;   2000US-0246523.
PR   08-NOV-2000;   2000US-0246524.
PR   08-NOV-2000;   2000US-0246525.
PR   08-NOV-2000;   2000US-0246526.
PR   08-NOV-2000;   2000US-0246527.
PR   08-NOV-2000;   2000US-0246528.
PR   08-NOV-2000;   2000US-0246532.
PR   08-NOV-2000;   2000US-0246609.
PR   08-NOV-2000;   2000US-0246610.
PR   08-NOV-2000;   2000US-0246611.
PR   08-NOV-2000;   2000US-0246613.
PR   17-NOV-2000;   2000US-0249207.
PR   17-NOV-2000;   2000US-0249208.
PR   17-NOV-2000;   2000US-0249209.
PR   17-NOV-2000;   2000US-0249210.
PR   17-NOV-2000;   2000US-0249211.
PR   17-NOV-2000;   2000US-0249212.
PR   17-NOV-2000;   2000US-0249213.
PR   17-NOV-2000;   2000US-0249214.
PR   17-NOV-2000;   2000US-0249215.
PR   17-NOV-2000;   2000US-0249216.
PR   17-NOV-2000;   2000US-0249217.
PR   17-NOV-2000;   2000US-0249218.
PR   17-NOV-2000;   2000US-0249244.
PR   17-NOV-2000;   2000US-0249245.
PR   17-NOV-2000;   2000US-0249264.
PR   17-NOV-2000;   2000US-0249265.
PR   17-NOV-2000;   2000US-0249297.
PR   17-NOV-2000;   2000US-0249299.
PR   17-NOV-2000;   2000US-0249300.
PR   01-DEC-2000;   2000US-0250160.
PR   01-DEC-2000;   2000US-0250391.
PR   05-DEC-2000;   2000US-0251030.
PR   05-DEC-2000;   2000US-0251988.
PR   05-DEC-2000;   2000US-0251989.
PR   06-DEC-2000;   2000US-0256719.
PR   06-DEC-2000;   2000US-0251479.
PR   08-DEC-2000;   2000US-0251856.
PR   08-DEC-2000;   2000US-0251868.
PR   08-DEC-2000;   2000US-0251869.
PR   08-DEC-2000;   2000US-0251990.
PR   11-DEC-2000;   2000US-0254097.
XX
PR   05-JAN-2001;   2001US-0259678.
```

PA   (HUMA-) HUMAN GENOME SCI INC.
XX
PI   Rosen CA,   Barash SC,   Ruben SM;
XX
DR   WPI; 2001-465570/50.
XX
PT   Disclosure; SEQ ID NO 9918; 1297pp + Sequence Listing; English.
XX
PS   Isolated nucleic acid molecule encoding a reproductive system antigen
PT   is used in preventing, treating or ameliorating a medical condition -
XX
CC   The present invention provides the protein and coding sequences of a
CC   number of human reproductive system related antigens. These can be used
CC   in the prevention and treatment of reproductive system disorders,
CC   including cancer. The present sequence is a genomic sequence encoding a
CC   protein of the invention.
XX
SQ   Sequence 5881 BP; 1445 A; 1534 C; 1610 G; 1292 T; 0 other;

      Query Match   25.1%; Score 250.8; DB 22; Length 5881;
      Best Local Similarity   88.1%; Pred. No. 5.3e-52;
      Matches 273; Conservative 0; Mismatches 37; Indels   0; Gaps   0;

Qy   466 ATTGCTGGCACGGTGGCTCATCCTGCTCAATCCACGCACCTTGGGAGGCTGAGGCAGGC 525
          |||||||||||||||||||||||||||| |||||| || |||||||||||||||||||
Db   3498 AATGGCTGGGCACGGTGGCTCACCGTGTAATCCCAGCACTTTGGGAGGCTGAGGTGGGT 3557

Qy   526 AGATGACTTGAGGCCAGGAGAGTTCGAGACCAGCCTGGCCAATATGGCGAAACCTCATTTCT 585
          || ||| |||||||||||||||||||||||||||||||||||||||||||| ||||||||
Db   3558 GGATCACCTGAGGTTGGGAGTTCGAGACCAGCCTGACCAAAATGGTGAAACCCCGTCTCT 3617

Qy   586 ACTAAAAATACAAAAAACTAACCAGGCGTGGTGGCTTGTGCCTGTAATCCCAGCTACTCGG 645
          |||||||| |||||||||| ||||||||||||||| ||||||||||||||||||||||||
Db   3618 ACTAAAAATTGCAAAAAATTAGCCCGGCGTGGTGGTGTGCGTGCCTGTAATCCCAGCTACTCAG 3677

Qy   646 GAGGCTGAGGCAGGAGAATCGCTTGAACCCGGGAGGCAGAGGTTGCAGTGAGCTGAGATC 705
          |||||||||||||||||||||||||||||||||||||||||||||||||| |||
Db   3678 GAGGCCGAGGCAGGAGAATCGCTTGAACCCGGGAGGCGGAGGTTACAGTGAGCCAAGATC 3737

Qy   706 ACACCACTGCACTCCAGCCTGGGCGACAGAGCGAGACTCCAGCTTTAAAAAAAAAAAAAA 765
          || || ||||||||||||||||||||||||||||||||| |||||||||||||
Db   3738 ACACCACTGCACTCCAGCCTGGGCGACAGAGCGAGACTCTGTCTCCAAAAAAAAAAAAAAA 3797

Qy   766 AAAAAAAAAA 775
          ||||||| ||
Db   3798 AAAAAAGAA 3807

RESULT 7
ABL98778
ID   ABL98778 standard; DNA; 5881 BP.
XX
AC   ABL98778;
XX
DT   21-JUN-2002   (first entry)
XX
DE   Human testicular antigen encoding DNA fragment SEQ ID NO: 3430.
XX
KW   Human; testicular antigen; testes; cancer; metastasis; immune disorder;
KW   reproductive system disorder; urinary system disorder; gene therapy;
KW   cardiovascular disorder; respiratory disorder; neurological disorder;
KW   gastrointestinal disease; infection; cytostatic; gene; ds.
XX
OS   Homo sapiens.
XX
PN   WO20015317-A2.
XX
PD   02-AUG-2001.
XX
PF   17-JAN-2001; 2001WO-US01329.
XX
PR   31-JAN-2000; 2001US-0179065.
PR   04-FEB-2000; 2000US-0180628.

PR 02-OCT-2000; 2000US-0237039.
PR 02-OCT-2000; 2000US-0237040.
PR 13-OCT-2000; 2000US-0239935.
PR 13-OCT-2000; 2000US-0239937.
PR 20-OCT-2000; 2000US-0240960.
PR 20-OCT-2000; 2000US-0241221.
PR 20-OCT-2000; 2000US-0241786.
PR 20-OCT-2000; 2000US-0241785.
PR 20-OCT-2000; 2000US-0241787.
PR 20-OCT-2000; 2000US-0241808.
PR 20-OCT-2000; 2000US-0241809.
PR 20-OCT-2000; 2000US-0241826.
PR 01-NOV-2000; 2000US-0244617.
PR 01-NOV-2000; 2000US-0244474.
PR 08-NOV-2000; 2000US-0246475.
PR 08-NOV-2000; 2000US-0246476.
PR 08-NOV-2000; 2000US-0246477.
PR 08-NOV-2000; 2000US-0246478.
PR 08-NOV-2000; 2000US-0246523.
PR 08-NOV-2000; 2000US-0246524.
PR 08-NOV-2000; 2000US-0246525.
PR 08-NOV-2000; 2000US-0246526.
PR 08-NOV-2000; 2000US-0246527.
PR 08-NOV-2000; 2000US-0246528.
PR 08-NOV-2000; 2000US-0246532.
PR 08-NOV-2000; 2000US-0246609.
PR 08-NOV-2000; 2000US-0246610.
PR 08-NOV-2000; 2000US-0246611.
PR 08-NOV-2000; 2000US-0246613.
PR 08-NOV-2000; 2000US-0249207.
PR 17-NOV-2000; 2000US-0249208.
PR 17-NOV-2000; 2000US-0249209.
PR 17-NOV-2000; 2000US-0249210.
PR 17-NOV-2000; 2000US-0249211.
PR 17-NOV-2000; 2000US-0249212.
PR 17-NOV-2000; 2000US-0249213.
PR 17-NOV-2000; 2000US-0249214.
PR 17-NOV-2000; 2000US-0249215.
PR 17-NOV-2000; 2000US-0249216.
PR 17-NOV-2000; 2000US-0249217.
PR 17-NOV-2000; 2000US-0249218.
PR 17-NOV-2000; 2000US-0249244.
PR 17-NOV-2000; 2000US-0249245.
PR 17-NOV-2000; 2000US-0249264.
PR 17-NOV-2000; 2000US-0249265.
PR 17-NOV-2000; 2000US-0249297.
PR 17-NOV-2000; 2000US-0249299.
PR 17-NOV-2000; 2000US-0249300.
PR 01-DEC-2000; 2000US-0250160.
PR 01-DEC-2000; 2000US-0250391.
PR 05-DEC-2000; 2000US-0251030.
PR 05-DEC-2000; 2000US-0251989.
PR 05-DEC-2000; 2000US-0251988.
PR 05-DEC-2000; 2000US-0256719.
PR 06-DEC-2000; 2000US-0251479.
PR 06-DEC-2000; 2000US-0251856.
PR 08-DEC-2000; 2000US-0251868.
PR 08-DEC-2000; 2000US-0251869.
PR 08-DEC-2000; 2000US-0251989.
PR 08-DEC-2000; 2000US-0251990.
PR 11-DEC-2000; 2000US-0254097.
PR 05-JAN-2001; 2000US-0259678.
XX
PA (HUMA-) HUMAN GENOME SCI INC.
XX
PI Rosen CA, Barash SC, Ruben SM;
XX
XX
PI WPI; 2001-483232/52.
XX
DR
PT Nucleic acids encoding 973 human testicular antigen polypeptides,
PT useful for preventing, diagnosing and/or treating testicular cancer -
XX
XX
PS Disclosure; SEQ ID NO 3430; 766pp; English.
XX

CC The present invention provides the protein and coding sequences of 973
CC human testicular antigens, and fragments of their genomic sequences. The
CC sequences can be used in the treatment of cardiovascular, urinary system,
CC reproductive system, immune, respiratory, neurological and
CC gastrointestinal disorders, infections, and particularly cancer,
CC especially testicular cancers. The present sequence is a DNA encoding a
CC protein fragment of the invention.
XX
SQ Sequence 5881 BP; 1445 A; 1534 C; 1610 G; 1292 T; 0 other;

Query Match 25.1%; Score 250.8; DB 23; Length 5881;
Best Local Similarity 88.1%; Pred No. 5.3e-52;
Matches 273; Conservative 0; Mismatches 37; Indels 0; Gaps 0;

```
Qy  466 ATTGGCTGGGCACGGTGGCCTCATGCCTGCCAATCCCAGCACCTTGGGAGGCTGAGGCAGGC 525
        ||||||||||||||||||||  |||||||| | |||||||||||||| ||||||||||||
Db 3498 AATGGCTGGGCACGGTGGCTCACGCCTGTAATCCCAGCACTTTGGGAGGCTGAGGTGGGT 3557

Qy  526 AGATGACTTGAGGCCAGGAGTTCGAGACCAGCCTGGCCAATATGGCGGAAACCTCATTTCT 585
        ||| || ||||||| +  |||||||||||||||| || ||||| |  ||||| |  ||
Db 3558 GGATCACCTGAGGTTGGAAGTTCGAGACCAGCCTGACCAAAATGGTGAAACCCCGTCTCT 3617

Qy  586 ACTAAAAATACAAAAAACTAACCAGGCGTGGTGGCGTGTGCCTGTAATCCCAGCTACTCGG 645
        |||||||| | |||| | |||||||||||||||| || ||||||| ||||||||| |||||
Db 3618 ACTAAAAATTGCAAAAAATTAGCCCGGCGTGGTGGCGGGCGCCTGTAATCCCAGCTACTCAG 3677

Qy  646 GAGGCTGAGGCAGGAGAATCGCTTGAACCGGGAGGCGGAGGGTTGCAGTGAGCTGAGATC 705
        ||||| |||||||| |||||||||||||||| ||||||||| |  ||||||| ||||||
Db 3678 GAGGCCGGAGGCAGAGAATCGCTTGAACCGGGGAGGCGGAGGTTACAGTGAGCCAAGATC 3737

Qy  706 ACACCACTGCACTCCAGCCTGGGCGACAGAGCGACTCCAGCTTAAAAAAAAAAAAAAA 765
        |||||||||||||||||||||||||||||||||||  |  |||||||||||||||||
Db 3738 ACACCACTGCACTCCAGCCTGGGCGACAGAGCGAGACTCTGTCTCAAAAAAAAAAAAAAA 3797

Qy  766 AAAAAAAAAA 775
        |||||||| ||
Db 3798 AAAAAAAGAA 3807
```

RESULT 8
ABN96830/c
ID ABN96830 standard; DNA; 40433 BP.
XX
AC ABN96830;
XX
DT 13-AUG-2002 (first entry)
XX
DE Gene #3328 used to diagnose liver cancer.
XX
OS Homo sapiens.
XX
KW Gene; liver cancer; dsj; hepatocellular carcinoma; hepatotropic;
KW metastatic liver tumour; cytostatic; expression profile; disease state;
KW disease progression; drug toxicity; drug efficacy; drug metabolism.
XX
PN WO20022910 3-A2.
XX
PD 11-APR-2002.
XX
PF 02-OCT-2001; 2001WO-US30589.
XX
PR 02-OCT-2000; 2000WO-237054P.
XX
PA (GENE-) GENE LOGIC INC.
XX
PI Horne D, Alvares C, Peres-Da-Silva S, Vockley JG;
XX
DR WPI; 2002-426119/45.
XX
PT Diagnosing and detecting the progression of liver cancer,
PT hepatocellular carcinoma or metastatic liver tumor in a patient,
PT involves detecting the level of expression of two or more genes in a
PT liver tissue sample -

XX
PS   Claim 1; SEQ ID NO 3328; 298pp; English.
XX
CC   The invention relates to a novel method for diagnosing and detecting the
CC   progression of liver cancer, hepatocellular carcinoma or metastatic liver
CC   tumour in a patient, and differentiating metastatic liver cancer from
CC   hepatocellular carcinoma in a patient, involving detecting the level of
CC   expression of two or more genes represented in ABN93503-ABN97455 in a
CC   tissue sample. The method of the invention has hepatotropic, and
CC   cytostatic activity. The method is useful for diagnosing and detecting
CC   the progression of liver cancer, hepatocellular carcinoma and metastatic
CC   liver carcinoma in a patient. The method is useful for identifying
CC   markers that can be used to monitor disease states, disease progression,
CC   expression profiles which serve as useful diagnostic markers as well as
CC   drug toxicity, drug efficacy and drug metabolism.
CC   Note: The sequence data for this patent did not form part of the printed
CC   specification, but was obtained in electronic format directly from WIPO
CC   at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ   Sequence 40433 BP; 11410 A; 8109 C; 8158 G; 12756 T; 0 other;

Query Match   25.0%; Score 250.2; DB 24; Length 40433;
Best Local Similarity   77.5%; Pred. No. 1.3e-51;
Matches 303; Conservative   0; Mismatches   88; Indels   0; Gaps   0;

Qy   466 ATTGGCGTGGGCACGGTGGCTCATGCCTGTAATCCCAGCACACTTTGGGAGGCCGAGGCAGGC 525
           ||||||||||||||||||||| ||||||||||||| |||||||||||||||||||||||||||
Db   9505 ACTGGCTGGCGACGGTGGGTCAATGCCTGCCAATCCTAGCACTTCGGGAGGCCGAGGTGGGC 9446

Qy   526 AGATGACTTGAGGCCAGGAGTTCGAGACCAGCCTGGCCAATATGGCGAAACCCCATTTTCT 585
           |||| ||||||||||||||||||||||||||||||||||||||||||| |||||||| |||
Db   9445 AGATCACTTGAGGTCAGGAGTTTGGGACCAGCCTGATCAATATGGTGAAAACCTTGTCTCT 9386

Qy   586 ACTAAAAATACAAAAAACTAACCAGGCGTGGTGGGTCGCTGTGTAATCCCAGCTACTCGG 645
           |||||||||||||||||||||||||||| ||||||||||||||| ||||||||||||||||
Db   9385 ACTAAAAATACAAAAATTAGCCAGGCATGGTGGTGCGCGCCTGTAGTCCCAGCTACTCAAG 9326

Qy   646 GAGGCTGAGGCAGGAGAAATCGCTTGAACCCGGGGAGGCGGAGGTTGCAGTGAGCTGAGATC 705
           |||||||||||||||||||||||||| ||||||||||||||||||||||||||||||||||
Db   9325 GGGGGCTGAGGCAGGAGAATCGCTTGAACCTGGGAGGCGGAGGTTGCAGTGAGCTGAGATC 9266

Qy   706 ACACCACTGCACTCCAGCCTGGGCGACAGAGCGAGACTCCAGCTTAAAAAAAAAAAAA 765
           |||||||||||||||||||||||||||||||||||||| |||||||||||||||||||||
Db   9265 GTGCCACTGCACTCCAGCCTGGGCAACAAAAGCTAGATTCCATCTCAAAAAAAAAAAAAACA 9206

Qy   766 AAAAAAAAAAAGGAGACCATCACTCGCTGTCCTGCCATTCTTACAGATGAAAAAACAGGCTCA 825
           |||| ||||||||||||||||||||||||||||||| ||||||||||||||||||||||||
Db   9205 AAAAACAAAAAAAAAGAAAAAAATGAAAAACTACAAGTGGGGAAGCCAGTTAG 9146

Qy   826 GAGGTTGAATCGTTTTCCTGAAGTCAGACAG 856
           |||||| |||| | | |||||||| | |||
Db   9145 GAAAGTTATATGATTAGTCCAAATGAGAGGCAG 9115

XX
PT   Detecting granulocyte activation by detecting differential expression
PT   of genes associated with granulocyte activation, which serves as
PT   diagnostic markers that is useful for monitoring disease states and
PT   drug toxicity.
XX
PS   Claim 1; SEQ ID NO 68; 114pp; English.
XX
CC   The invention relates to detecting (M1) granulocyte (GC) activation
CC   (GCA), by detecting the level of expression of gene(s) (Gs) identified by
CC   DNA chip analysis as given in the specification, and comparing the
CC   expression level to an expression level in an unactivated
CC   GC, where differential expression of Gs is indicative of GCA.
CC   Also included are modulating GC by contacting GC with an agent
CC   that alters the expression of at least one gene in Gs; (2) screening (M3)
CC   for an agent capable of modulating GCA or an inflammation (especially
CC   chronic) in a tissue, an allergic response in a subject, exposure of a
CC   subject to a pathogen or sterile inflammatory disease using the
CC   gene expression profile; (3) detecting (M4) an inflammation (especially
CC   chronic) in a tissue, an allergic response in a subject, exposure of a
CC   subject to a pathogen or sterile inflammatory disease, by detecting the
CC   level of expression of the gene(s) in a sample of the tissue of gene(s)
CC   (4) treating (M5) an inflammation (especially chronic) or in a tissue,
CC   an allergic response in a subject, exposure of a subject to a pathogen
CC   or sterile inflammatory disease, by contacting a tissue having
CC   inflammation with an agent that modulates the expression of gene(s)
CC   from Gs in the tissue. M1 is useful for detecting GCA. M2 is useful for
CC   modulating GA; M3 is useful for screening an agent capable of modulating
CC   GCA preferably in an inflammation (especially chronic) in a tissue, an
CC   detecting an inflammation (especially chronic) in a tissue, an allergic
CC   response in a subject, exposure of a subject to a pathogen or sterile
CC   inflammatory disease (e.g. psoriasis, rheumatoid arthritis,
CC   glomerulonephritis, asthma, thrombosis, cardiac reperfusion injury, renal
CC   reperfusion injury, ARDS, adult respiratory distress syndrome,
CC   inflammatory bowel disease, Crohn's disease, ulcerative colitis,
CC   periodontal disease; also bacterial infection, fungal infection, viral
CC   parasitic infection, protozoal infection, fungal infection and M5 is
CC   useful for treating one of the above conditions. The present
CC   sequence represents a gene differentially expressed in granulocytes.
CC   Note: The sequence data for this patent did not form part
CC   of the printed specification, but was obtained in electronic
CC   format directly from WIPO at
CC   ftp.wipo.int/pub/published_pct_sequences.
XX
SQ   Sequence 227968 BP; 53185 A; 60440 C; 61985 G; 52358 T; 0 other;

XX
DE   Human cDNA differentially expressed in granulocytic cells #68.
XX
AC   ABK83497;
XX
DT   14-AUG-2002 (first entry)

RESULT 9
ID   ABK83497/c
ID   ABK83497 standard; cDNA; 227968 BP.
XX
AC   ABK83497;
XX
DT   14-AUG-2002 (first entry)
XX
DE   Human cDNA differentially expressed in granulocytic cells #68.
XX
KW   granulocytic cell; DNA chip; bacterial infection;
KW   viral infection; parasitic infection; protozoal infection;
KW   fungal infection; sterile inflammatory disease; psoriasis;
KW   rheumatoid arthritis; glomerulonephritis; asthma; thrombosis;
KW   cardiac reperfusion injury; renal reperfusion injury; ARDS;
KW   adult respiratory distress syndrome; inflammatory bowel disease;
KW   Crohn's disease; ulcerative colitis; periodontal disease;
KW   granulocyte activation; chronic inflammation; allergy.
XX
OS   Human; ss; granulocytic cell.

Query Match   24.9%; Score 249.6; DB 24; Length 227968;
Best Local Similarity   86.2%; Pred. No. 3e-51;
Matches 276; Conservative   0; Mismatches   44; Indels   0; Gaps   0;

Qy   456 AGAGACAGAGTCATTTGGCTGGGCACGGTGGCTCATGCCTGTAATCCCAGCACCTTTGGGAGGC 515
           || |||||||||||||||||| |||||||||||||||||||||||||||||||||||||||||
Db   30928 AAAGACAGATATTGGCTGGGTGCATGGCTCATGCCTGTAATCCTAGCACTTTGGGAGGC 30869

Qy   516 TGAGGCAGGCAGATGACTTGAGGCCAGGAGTTCGAGACCAGCCTGGCCAATATGGCGAAA 575
           |||||| ||||| |||||||||||||||||||||||||||||||||||||||||||| |||
Db   30868 CAAGGCGGCGCGGATCACCTGAGGTCAGGAGTTCGAGACCAGCCTGGCCAACATGGCGAAA 30809

```
Qy     576 CCTCATTTCTTACTAAAAATACAAAAACTAACCAGGCGTGGTGGCTTGTGCCTGTAATCCC 635
           || |||||||| |||||||||||||||||||| ||| |||||||||||||||||||||||||
Db   30808 CCAAGTCTCTACTAAAAATACAAAAATTAGCCGGGCGTGGTGGCGCGCGCCTGTAATCCC 30749

Qy     636 AGCTACTCGGGAGGCTGAGGCAGGAGAATCCCTTGAACCGGGGAGGCAGAGGTTGCAGTG 695
           ||||||||||||||||||||||||||||||||| || |||||||||||||||||||||||||
Db   30748 AGCTACTCTGGAGGCTGAGGCAGGAGAATCCGTTGAACGCGCTGAAGGCGGGAGGTTGCAGTG 30689

Qy     696 AGCTGAGATCACACCACTGCACTCCAGCCTGGGCGACGAGGAGACTCCAGCTTAAAAA 755
           ||||||| |||||| ||||||||||||||||||||||||||||| ||||||||||||||
Db   30688 AGCCAAGATCACGCCACTTCACTCCAGCCTGGGTGACAGAGTGAGACTCCGTCTCAAAAA 30629

Qy     756 AAAAAAAAAAAAAAAA 775
           |||||||||| ||||||
Db   30628 AAAAAAAAAAACAAAAAA 30609
```

RESULT 10
ABA17153

ID   ABA17153 standard; DNA; 5041 BP.
XX
AC   ABA17153;
XX
DT   23-JAN-2002 (first entry)
XX
DE   Human nervous system related polynucleotide SEQ ID NO 9484.
XX
KW   Human; nootropic; neuroprotective; cytostatic; dermatological; virucide;
KW   immunosuppressive; antiinflammatory; anti-HIV; antibacterial; vulnerary;
KW   antiparkinsonian; antisickling; antianaemic; antiarthritic; cancer;
KW   antirheumatic; hepatotropic; cerebroprotective; antiinflammatory;
KW   antiallergic; antidiabetic; antiulcer; anticonvulsant; antifungal;
KW   antiparasitic; cardiant; immune disorder; cardiovascular disorder;
KW   neurological disease; infection; nephrotropic; gene therapy; vaccine; ds.
XX
OS   Homo sapiens.
XX
PN   WO200159063-A2.
XX
PD   16-AUG-2001.
XX
PF   17-JAN-2001; 2001WO-US01334.
XX
PR   31-JAN-2000; 2000US-0179065.
PR   04-FEB-2000; 2000US-0180628.
PR   24-FEB-2000; 2000US-0184664.
PR   02-MAR-2000; 2000US-0186350.
PR   16-MAR-2000; 2000US-0189074.
PR   17-MAR-2000; 2000US-0190076.
PR   18-APR-2000; 2000US-0198123.
PR   19-MAY-2000; 2000US-0205515.
PR   07-JUN-2000; 2000US-0209467.
PR   28-JUN-2000; 2000US-0214886.
PR   30-JUN-2000; 2000US-0215135.
PR   26-JUL-2000; 2000US-0220964.
PR   14-AUG-2000; 2000US-0220963.
PR   14-JUL-2000; 2000US-0218290.
PR   11-JUL-2000; 2000US-0217496.
PR   11-JUL-2000; 2000US-0217487.
PR   07-JUL-2000; 2000US-0216880.
PR   07-JUL-2000; 2000US-0216647.
PR   20-JUL-2000; 2000US-0219513.
PR   06-SEP-2000; 2000US-0230437.
PR   06-SEP-2000; 2000US-0230438.
PR   06-SEP-2000; 2000US-0231242.
PR   08-SEP-2000; 2000US-0231243.
PR   08-SEP-2000; 2000US-0231244.
PR   08-SEP-2000; 2000US-0231413.
PR   08-SEP-2000; 2000US-0231414.
PR   08-SEP-2000; 2000US-0232080.
PR   08-SEP-2000; 2000US-0232081.
PR   12-SEP-2000; 2000US-0231968.
PR   14-SEP-2000; 2000US-0232397.
PR   14-SEP-2000; 2000US-0232398.
PR   14-SEP-2000; 2000US-0232399.
PR   14-SEP-2000; 2000US-0232400.
PR   14-SEP-2000; 2000US-0232401.
PR   14-SEP-2000; 2000US-0233063.
PR   14-SEP-2000; 2000US-0233064.
PR   14-SEP-2000; 2000US-0233065.
PR   21-SEP-2000; 2000US-0234223.
PR   21-SEP-2000; 2000US-0234224.
PR   21-SEP-2000; 2000US-0234274.
PR   25-SEP-2000; 2000US-0234997.
PR   25-SEP-2000; 2000US-0234998.
PR   26-SEP-2000; 2000US-0235484.
PR   27-SEP-2000; 2000US-0235834.
PR   27-SEP-2000; 2000US-0235836.
PR   29-SEP-2000; 2000US-0236327.
PR   29-SEP-2000; 2000US-0236367.
PR   29-SEP-2000; 2000US-0236368.
PR   29-SEP-2000; 2000US-0236369.
PR   02-OCT-2000; 2000US-0236370.
PR   02-OCT-2000; 2000US-0236802.
PR   02-OCT-2000; 2000US-0237037.
PR   02-OCT-2000; 2000US-0237038.
PR   02-OCT-2000; 2000US-0237039.
PR   02-OCT-2000; 2000US-0237040.
PR   13-OCT-2000; 2000US-0239935.
PR   13-OCT-2000; 2000US-0239937.
PR   20-OCT-2000; 2000US-0240960.
PR   20-OCT-2000; 2000US-0241785.
PR   20-OCT-2000; 2000US-0241786.
PR   20-OCT-2000; 2000US-0241787.
PR   20-OCT-2000; 2000US-0241808.
PR   20-OCT-2000; 2000US-0241809.
PR   20-OCT-2000; 2000US-0241826.
PR   01-NOV-2000; 2000US-0244617.
PR   08-NOV-2000; 2000US-0242221.
PR   08-NOV-2000; 2000US-0246474.
PR   08-NOV-2000; 2000US-0246475.
PR   08-NOV-2000; 2000US-0246476.
PR   08-NOV-2000; 2000US-0246477.
PR   08-NOV-2000; 2000US-0246478.
PR   08-NOV-2000; 2000US-0246523.
PR   08-NOV-2000; 2000US-0246524.
PR   08-NOV-2000; 2000US-0246525.
PR   08-NOV-2000; 2000US-0246526.
PR   08-NOV-2000; 2000US-0246527.
PR   08-NOV-2000; 2000US-0246528.
PR   08-NOV-2000; 2000US-0246532.
PR   08-NOV-2000; 2000US-0246609.
PR   08-NOV-2000; 2000US-0246610.
```

And in the first preference-number column:
```
PR   14-AUG-2000; 2000US-0225759.
PR   18-AUG-2000; 2000US-0226279.
PR   22-AUG-2000; 2000US-0226681.
PR   22-AUG-2000; 2000US-0226868.
PR   22-AUG-2000; 2000US-0227182.
PR   23-AUG-2000; 2000US-0227009.
PR   30-AUG-2000; 2000US-0228924.
PR   01-SEP-2000; 2000US-0229287.
PR   01-SEP-2000; 2000US-0229343.
PR   01-SEP-2000; 2000US-0229344.
PR   01-SEP-2000; 2000US-0229345.
PR   05-SEP-2000; 2000US-0229513.
PR   14-AUG-2000; 2000US-0225758.
PR   14-AUG-2000; 2000US-0225757.
PR   14-AUG-2000; 2000US-0225447.
PR   14-AUG-2000; 2000US-0225270.
PR   14-AUG-2000; 2000US-0225268.
PR   14-AUG-2000; 2000US-0225267.
PR   14-AUG-2000; 2000US-0225266.
PR   14-AUG-2000; 2000US-0225214.
PR   14-AUG-2000; 2000US-0225213.
PR   14-AUG-2000; 2000US-0224519.
PR   14-AUG-2000; 2000US-0224518.
```

```
PR   08-NOV-2000;  2000US-0246611.
PR   08-NOV-2000;  2000US-0246613.
PR   17-NOV-2000;  2000US-0249207.
PR   17-NOV-2000;  2000US-0249208.
PR   17-NOV-2000;  2000US-0249209.
PR   17-NOV-2000;  2000US-0249210.
PR   17-NOV-2000;  2000US-0249211.
PR   17-NOV-2000;  2000US-0249212.
PR   17-NOV-2000;  2000US-0249213.
PR   17-NOV-2000;  2000US-0249214.
PR   17-NOV-2000;  2000US-0249215.
PR   17-NOV-2000;  2000US-0249216.
PR   17-NOV-2000;  2000US-0249217.
PR   17-NOV-2000;  2000US-0249218.
PR   17-NOV-2000;  2000US-0249244.
PR   17-NOV-2000;  2000US-0249245.
PR   08-DEC-2000;  2000US-0251856.
PR   08-DEC-2000;  2000US-0251868.
PR   08-DEC-2000;  2000US-0251869.
PR   08-DEC-2000;  2000US-0251989.
PR   08-DEC-2000;  2000US-0251990.
PR   11-DEC-2000;  2000US-0254097.
PR   11-DEC-2000;  2000US-0254097.
PR   05-JAN-2001;  2001US-0259678.
```

```
Db    2986  CTCTTAAAAGAACCTTTCCAGCCGCGCAGTGGCTCATGGCCTGTAATCCCAGCACTTTG  3045
            ||||  |||  ||||  ||  |||  ||||||||||||||  |||||||||||||||||||||||  |||
Db    510  GGAGGCTGAGGCAGGCAGAGTGAGACTTGAGGCCAGGCCAGAGGAGTTCGAGACCAGCCTGGCCAATATG  569
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy    3046  GGAGGCCGAGGCAGGCAGAGATGAGTGACTTGAGGTCGAGAGTTCGAGACCAGCCTGGCCAACGTG  3105
            ||||||||||||||||||||||||  |||  ||||  |||||  ||||||||||||||  ||||||  |||||
Qy    570  GCGAAACCTCATTTCTACTAAAAATACAAAAAAACTAACCAGGCGTGGTGGCTGTGGCCTGT  629
            ||  ||||||||||  ||||||||||||  ||||  ||||||||  |||  ||||||||  ||||||  ||||
Db    3106  GCAAAACCCCATCTCTACTAAAAATAAGAAAATTAGCCGGGTGTGGTGGCACATGCCTGT  3165
Qy    630  AATCCCAGCTACTCGGGAGGCTGAGGCAGGAGAATCGCTTGAACCCGGGAGGCGGAGGTT  689
            ||||||||||||||||  ||||||||||||||||  |||  ||||||||||||  ||||||||||
Db    3166  AATCCCAGCTACTCAGGAGGCTGAGGCAGGAGAATCGCTTGAGCCCAGGAGGCGGAGGATT  3225
Qy    690  GCAGTGAGCTGAGATCACACACCACTGCACTCCAGCCTGGGCGACAGAGCGAGACTCCAGCT  749
            |||||||||||||||  ||||||||  ||||||  |||||||||  |||||  |||||||||||||||
Db    3226  GCAGTGAGCCAAGATCGCCACCACTGCACTCCAGCCTCGGTGACAGAGCCAAGACTCCATCT  3285
Qy    750  TAAAAAAAAAAAAAAAAAAAAAAAAA  772
            |||||||||||||||||||||||||
Db    3286  CAGAAAAAAAAAAAAAAAAAAAAAAA  3308
```

```
PI   Rosen CA,  Barash SC,  Ruben SM;
PA   (HUMA-) HUMAN GENOME SCI INC.

DR   WPI; 2001-541565/60.

PT   Nucleic acids encoding 3224 human nervous system antigen polypeptides,
PT   useful for preventing, diagnosing and/or treating nervous system
XX   cancers and metastases -

PS   Disclosure; SEQ ID NO 9484; 1701pp + Sequence Listing; English.

CC   The invention relates to novel genes (ABA11004-ABA21534) and proteins
CC   (ABB14678-ABB18001) useful for preventing, treating or ameliorating
CC   medical conditions e.g. by protein or gene therapy. The genes are
CC   isolated from a range of human tissues disclosed in the specification.
CC   The nucleic acids, proteins, antibodies and (ant)agonists are useful
CC   in the diagnosis, treatment and prevention of: (a) cancer, e.g. breast
CC   and ovarian cancer and other cancers of the adrenal gland, bone, bone
CC   marrow, breast, gastrointestinal tract, liver, lung, or urogenital;
CC   (b) immune disorders e.g. Addison's disease, allergies, autoimmune
CC   haemolytic anaemia, autoimmune thyroiditis, diabetes mellitus, Crohn's
CC   disease, multiple sclerosis, rheumatoid arthritis and ulcerative
CC   colitis; (c) cardiovascular disorders such as myocardial ischaemias;
CC   (d) wound healing; (e) neurological diseases e.g. cerebral anoxia and
CC   epilepsy; and (f) infectious diseases such as viral, bacterial, fungal
CC   and parasitic infections.
CC   Note: The sequence data for this patent did not form part of the
CC   printed specification, but was obtained in electronic format directly
CC   from WIPO at ftp.wipo.int/pub/published_pct_sequences.

SQ   Sequence 5041 BP; 1480 A; 1074 C; 1103 G; 1384 T; 0 other;
```

```
Qy   450  CTCTCAAGAGACAGTCATTGGCTGTGGGCACGGTGGCTCATGCCTGCGTAATCCCAGCACCTTG  509
```

Query Match   24.9%;   Score 249.4;   DB 22;   Length 5041;
Best Local Similarity   85.8%;   Pred. No. 1.1e-51;
Matches 277;   Conservative 0;   Mismatches 46;   Indels 0;   Gaps 0;

```
RESULT 11
AAK89230/c
ID   AAK89230 standard; DNA; 30110 BP.
XX
AC   AAK89230;
XX
DT   05-NOV-2001  (first entry)
XX
DE   Human digestive system antigen genomic sequence SEQ ID NO: 2806.
XX
KW   Human; digestive system antigen; gene therapy; cancer; appendicitis;
KW   ulcerative colitis; infection; Hirschsprung's disease; chronic colitis;
KW   digestive system disorder; Meckel's diverticulum; ds.
XX
OS   Homo sapiens.
XX
PN   WO200155314-A2.
XX
PD   02-AUG-2001.
XX
PF   17-JAN-2001;  2001WO-US01324.
XX
PR   31-JAN-2000;  2000US-0179065.
PR   04-FEB-2000;  2000US-0180628.
PR   24-FEB-2000;  2000US-0184664.
PR   02-MAR-2000;  2000US-0184650.
PR   16-MAR-2000;  2000US-0189874.
PR   17-MAR-2000;  2000US-0190076.
PR   18-APR-2000;  2000US-0198123.
PR   19-MAY-2000;  2000US-0205515.
PR   07-JUN-2000;  2000US-0209467.
PR   28-JUN-2000;  2000US-0214886.
PR   30-JUN-2000;  2000US-0215135.
PR   07-JUL-2000;  2000US-0216647.
PR   07-JUL-2000;  2000US-0216880.
PR   11-JUL-2000;  2000US-0217487.
PR   11-JUL-2000;  2000US-0217496.
PR   14-JUL-2000;  2000US-0218290.
PR   26-JUL-2000;  2000US-0220963.
PR   26-JUL-2000;  2000US-0220964.
PR   14-AUG-2000;  2000US-0224518.
PR   14-AUG-2000;  2000US-0224519.
PR   14-AUG-2000;  2000US-0225213.
PR   14-AUG-2000;  2000US-0225214.
PR   14-AUG-2000;  2000US-0225266.
PR   14-AUG-2000;  2000US-0225267.
PR   14-AUG-2000;  2000US-0225268.
PR   14-AUG-2000;  2000US-0225270.
PR   14-AUG-2000;  2000US-0225447.
```

```
PR   14-AUG-2000;   2000US-0225757.
PR   14-AUG-2000;   2000US-0225758.
PR   14-AUG-2000;   2000US-0225759.
PR   18-AUG-2000;   2000US-0226279.
PR   22-AUG-2000;   2000US-0226681.
PR   22-AUG-2000;   2000US-0226868.
PR   22-AUG-2000;   2000US-0227182.
PR   23-AUG-2000;   2000US-0227009.
PR   30-AUG-2000;   2000US-0228924.
PR   01-SEP-2000;   2000US-0229287.
PR   01-SEP-2000;   2000US-0229343.
PR   01-SEP-2000;   2000US-0229344.
PR   01-SEP-2000;   2000US-0229345.
PR   05-SEP-2000;   2000US-0229509.
PR   05-SEP-2000;   2000US-0229513.
PR   06-SEP-2000;   2000US-0230437.
PR   06-SEP-2000;   2000US-0230438.
PR   08-SEP-2000;   2000US-0231242.
PR   08-SEP-2000;   2000US-0231243.
PR   08-SEP-2000;   2000US-0231244.
PR   08-SEP-2000;   2000US-0231413.
PR   08-SEP-2000;   2000US-0231414.
PR   08-SEP-2000;   2000US-0233063.
PR   08-SEP-2000;   2000US-0233064.
PR   08-SEP-2000;   2000US-0233065.
PR   08-SEP-2000;   2000US-0232080.
PR   08-SEP-2000;   2000US-0232081.
PR   08-SEP-2000;   2000US-0234223.
PR   12-SEP-2000;   2000US-0231968.
PR   14-SEP-2000;   2000US-0232397.
PR   14-SEP-2000;   2000US-0232398.
PR   14-SEP-2000;   2000US-0232399.
PR   14-SEP-2000;   2000US-0232400.
PR   14-SEP-2000;   2000US-0232401.
PR   14-SEP-2000;   2000US-0234274.
PR   14-SEP-2000;   2000US-0234997.
PR   14-SEP-2000;   2000US-0234998.
PR   21-SEP-2000;   2000US-0234997.
PR   25-SEP-2000;   2000US-0234998.
PR   25-SEP-2000;   2000US-0235484.
PR   26-SEP-2000;   2000US-0235834.
PR   27-SEP-2000;   2000US-0235836.
PR   27-SEP-2000;   2000US-0236327.
PR   29-SEP-2000;   2000US-0236367.
PR   29-SEP-2000;   2000US-0236368.
PR   29-SEP-2000;   2000US-0236369.
PR   29-SEP-2000;   2000US-0236370.
PR   29-SEP-2000;   2000US-0236802.
PR   02-OCT-2000;   2000US-0237037.
PR   02-OCT-2000;   2000US-0237038.
PR   02-OCT-2000;   2000US-0237039.
PR   02-OCT-2000;   2000US-0237040.
PR   13-OCT-2000;   2000US-0239935.
PR   13-OCT-2000;   2000US-0239937.
PR   02-OCT-2000;   2000US-0240960.
PR   02-OCT-2000;   2000US-0241221.
PR   20-OCT-2000;   2000US-0241785.
PR   20-OCT-2000;   2000US-0241786.
PR   20-OCT-2000;   2000US-0241787.
PR   20-OCT-2000;   2000US-0241808.
PR   20-OCT-2000;   2000US-0241809.
PR   20-OCT-2000;   2000US-0241826.
PR   01-NOV-2000;   2000US-0244617.
PR   08-NOV-2000;   2000US-0246474.
PR   08-NOV-2000;   2000US-0246475.
PR   08-NOV-2000;   2000US-0246476.
PR   08-NOV-2000;   2000US-0246477.
PR   08-NOV-2000;   2000US-0246478.
PR   08-NOV-2000;   2000US-0246523.
PR   08-NOV-2000;   2000US-0246524.
PR   08-NOV-2000;   2000US-0246525.
PR   08-NOV-2000;   2000US-0246526.
PR   08-NOV-2000;   2000US-0246527.
PR   08-NOV-2000;   2000US-0246528.
PR   08-NOV-2000;   2000US-0246532.

PR   08-NOV-2000;   2000US-0246609.
PR   08-NOV-2000;   2000US-0246610.
PR   08-NOV-2000;   2000US-0246611.
PR   08-NOV-2000;   2000US-0246613.
PR   17-NOV-2000;   2000US-0249207.
PR   17-NOV-2000;   2000US-0249208.
PR   17-NOV-2000;   2000US-0249209.
PR   17-NOV-2000;   2000US-0249210.
PR   17-NOV-2000;   2000US-0249211.
PR   17-NOV-2000;   2000US-0249212.
PR   17-NOV-2000;   2000US-0249213.
PR   17-NOV-2000;   2000US-0249214.
PR   17-NOV-2000;   2000US-0249215.
PR   17-NOV-2000;   2000US-0249216.
PR   17-NOV-2000;   2000US-0249217.
PR   17-NOV-2000;   2000US-0249218.
PR   17-NOV-2000;   2000US-0249244.
PR   17-NOV-2000;   2000US-0249245.
PR   17-NOV-2000;   2000US-0249264.
PR   17-NOV-2000;   2000US-0249265.
PR   17-NOV-2000;   2000US-0249297.
PR   17-NOV-2000;   2000US-0249299.
PR   17-NOV-2000;   2000US-0249300.
PR   01-DEC-2000;   2000US-0250160.
PR   01-DEC-2000;   2000US-0250391.
PR   05-DEC-2000;   2000US-0251030.
PR   05-DEC-2000;   2000US-0251988.
PR   05-DEC-2000;   2000US-0256719.
PR   06-DEC-2000;   2000US-0251479.
PR   08-DEC-2000;   2000US-0251856.
PR   08-DEC-2000;   2000US-0251868.
PR   08-DEC-2000;   2000US-0251869.
PR   08-DEC-2000;   2000US-0251989.
PR   08-DEC-2000;   2000US-0251990.
PR   11-DEC-2000;   2000US-0254097.
PR   05-JAN-2001;   2001US-0259678.
```

XX
PA (HUMA-) HUMAN GENOME SCI INC.
XX
PI Rosen CA, Barash SC, Ruben SM;
XX
DR WPI; 2001-502630/55.
XX
PT Polynucleotides encoding digestive system antigens, useful for
PT diagnosing, treating, preventing and/or prognozing disorders of the
PT digestive system, particularly cancer and cancer metastases -
XX
PS Disclosure; SEQ ID NO 2806; 986pp; English.
XX
CC The present invention provides the protein and coding sequences of a
CC number of human digestive system antigens. These can be used in the
CC diagnosis, treatment and prevention of digestive system disorders,
CC including cancer, Meckel's diverticulum, bacterial or parasitic
CC infections, appendicitis, Hirschsprung's disease, chronic colitis or
CC ulcerative colitis. The present sequence is a genomic DNA fragment
CC encoding a digestive system antigen of the invention.
XX
SQ Sequence 30110 BP; 7884 A; 6779 C; 6693 G; 8754 T; 0 other;

```
Query Match             24.8%;  Score 248.2;  DB 22;  Length 30110;
Best Local Similarity   84.1%;
Matches  280;  Conservative    0;  Mismatches   53;  Indels    0;  Gaps    0;
                                      Pred. No. 3.7e-51;

Qy    453 TCAAGAGACAGTCATTGGCTGGGCACGGTGGCTCATGCCTGCAATCCCAGCACCTTGGGA 512
          | |||| |  |||       || ||||||||||||||||||  ||||| |  ||| |||||
Db   3592 TTAAGATAGTAGCACAGGCTGGACACAGTGGCTCATGCCTGTAATCCCAACATTTTGGGA 3533

Qy    513 GGCTGAGGCAGGCAGAGATGACTTGAGGCCAGGAGTTCGAGACCAGCCTGGCCAATATGCG 572
          ||| |||||||| | ||||  |||||||| ||||||||||||||||||| |||||||  ||
Db   3532 GGCCGAGGCAGGACGGATCACTTGAGGTTCAGAGACCAGCCCTGGCCAACATGATG 3473

Qy    573 AAACCTCATTTCTACTAAAAATACAAAAACTAACCAGGGGTGGTGGCTTGTGCCTGTAAT 632
          |||||||||||  |  ||||||| ||||  ||||||||| |  ||||| | |||||||||
```

Db   3472 AAACCCCATCTCTACTAAAAATACAAAAATTAGCTGGGCGTGGTGGTGCGCCACCTGTAAT 3413
XX
Qy   633 CCCAGCTACTCGGGAGGCTGAGGCAGGAGGAATCGGCTTGAACCGGGAGGCGGAGGGTTGCA 692
          ||| || |||||||||||||||||||||||||||||| |||||| |||| || |||||
Db   3412 CCTAGCTACTCGGGAGGCTGAGGCAGGAGAGAATCTCTTGAACCTGGGAGGCGGAGCTTGCA 3353
XX
Qy   693 GTGAGCTGAGATCACACCACCACTGCACTCCAGCCTGGGCGACAGAGCGAGACTCCAGCTTAA 752
          |||||||||||||| | ||||||||| |||| ||||||| |||||| ||||||||| |||
Db   3352 GTGAGCTGAGATCACGCCACTGCACTCTAGCCCTGGGCAACAGAGTGAGACTCCAYCTCAA 3293
XX
Db   3292 AAAAAAAAAAAAAAAAAAAAAAAAAGAGACCATC 3260
          ||||||||||||||||||||||| |||| |||
Qy   753 AAAAAAAAAAAAAAAAAAAAAAAAAAGAGACCATCAA 785

RESULT 12
AAS21306
ID   AAS21306 standard; cDNA; 4374 BP.
XX
AC   AAS21306;
XX
DT   24-OCT-2001 (first entry)
XX
DE   Human cDNA sequence encoding for PRO4350 polypeptide.
XX
KW   Human secretory and transmembrane; PRO; mammalian; cancer; lung;
KW   breast; prostate; cervical; tumour necrosis factor-alpha; TNF-alpha;
KW   cartilage; ear; proliferation; glucose; free fatty acid; skeletal muscle;
KW   adipocyte; A-peptide; factor VIIA; gene therapy; ss.
XX
OS   Homo sapiens.
XX
XX
PN   WO200140466-A2.
XX
PD   07-JUN-2001.
XX
PF   01-DEC-2000; 2000WO-US32678.
XX
PR   01-DEC-1999; 99WO-US28301.
PR   01-DEC-1999; 99WO-US28634.
PR   02-DEC-1999; 99WO-US28551.
PR   02-DEC-1999; 99WO-US28564.
PR   02-DEC-1999; 99WO-US28565.
PR   09-DEC-1999; 99US-0170262.
PR   16-DEC-1999; 99WO-US30095.
PR   20-DEC-1999; 99WO-US30911.
PR   20-DEC-1999; 99WO-US30999.
PR   30-DEC-1999; 99WO-US31243.
PR   06-JAN-2000; 2000WO-US00277.
PR   06-JAN-2000; 2000WO-US00376.
PR   11-FEB-2000; 2000WO-US03565.
PR   18-FEB-2000; 2000WO-US04341.
PR   18-FEB-2000; 2000WO-US04342.
PR   22-FEB-2000; 2000WO-US04414.
PR   24-FEB-2000; 2000WO-US04914.
PR   24-FEB-2000; 2000WO-US05004.
PR   01-MAR-2000; 2000WO-US05601.
PR   20-MAR-2000; 2000WO-US07377.
PR   21-MAR-2000; 2000WO-US07532.
PR   30-MAR-2000; 2000WO-US08439.
PR   17-MAY-2000; 2000WO-US13705.
PR   22-MAY-2000; 2000WO-US14042.
PR   30-MAY-2000; 2000WO-US14941.
PR   02-JUN-2000; 2000WO-US15264.
PR   10-NOV-2000; 2000WO-US30873.
XX
PA   (GETH) GENENTECH INC.
XX
PI   Baker KP, Beresini M, Deforge L, Desnoyers L, Filvaroff E, Gao W;
PI   Gerritsen ME, Goddard A, Godowski PJ, Gurney AL, Sherwood S;
PI   Smith V, Stewart TA, Tumas D, Watanabe CK, Wood WI, Zhang Z;
XX
DR   WPI; 2001-408281/43.

DR   P-PSDB; AAU12234.
XX
PT   Isolated, secretory and transmembrane PRO polypeptide used to detect
PT   other PRO polypeptides, link bioactive molecules to cells expressing
PT   PRO polypeptides, and detect the presence of mammalian tumours e.g.
PT   lung, breast, prostate, cervical
XX
PS   Claim 3; Fig 125; 813pp; English.
XX
CC   AAS21244-AAS21518 encode for novel human secretory and transmembrane
CC   PRO polypeptides. The PRO polypeptides are useful to detect other
CC   PRO polypeptides, to link bioactive molecules to cells expressing
CC   PRO polypeptides, to modulate biological activities of cells expressing
CC   PRO polypeptides, and to detect the presence of mammalian lung, colon,
CC   breast, prostate, rectal, cervical or liver tumours by comparing PRO
CC   polypeptide expression in a cell sample to that in a control sample.
CC   Some of the 275 sequences are also useful to stimulate the release of
CC   tumour necrosis factor-alpha (TNF-alpha) from human blood, the
CC   proliferation or differentiation of chondrocytes, the proliferation or
CC   gene expression in pericyte cells, the release of proteoglycans from
CC   cartilage, the proliferation of inner ear utricular supporting cells or
CC   of T-lymphocytes, the release of a cytokine from peripheral blood
CC   monocytes (PBMCs), or the proliferation of endothelial cells. Some of
CC   the PRO polypeptides may modulate glucose or free fatty acid uptake by
CC   skeletal muscle cells or by adipocytes; or inhibit binding of A-peptide
CC   to factor VIIA. The PRO polypeptides can be used in assays to identify
CC   molecules involved in binding interactions. The polynucleotides encoding
CC   PRO polypeptides can be used to generate probes, antisense RNA/DNA,
CC   transgenic or knock out animals and can be used in gene therapy.
XX
SQ   Sequence 4374 BP; 1403 A; 824 C; 869 G; 1278 T; 0 other;
XX

Query Match   24.8%; Score 248; DB 22; Length 4374;
Best Local Similarity   76.2%;
Matches 305; Conservative   0; Mismatches   95; Indels   0; Gaps   0;

Qy   451 TCTCAAGAGACAGTCATTGGCTGGGGCACGGTGGCCTCATGCCTGCAATCCCAGCACCTTGG 510
          |||| | | | | | || | | | | | | || |||||||| | |||||||||||
Db   3112 TTTTAAAATATGTTACTCTGGCCGGGTTCCAGTGGCCTCACGCCCGTTAATCCCACCACTTGG 3171

Qy   511 GAGGCTGAGGCAGGCAGAGAGACTTGAGGCCAGGAGTTCGAGAACCAGCCCTGGCCAATATGG 570
          |||| ||| ||| || | || |||||| | |||||||||||||||||||||||||| |||||
Db   3172 CAGGCCGAGAGATGGGCGGGGATCTCTTTTGAGGTCAGGAGTTTGAGAACCAGCCTGGCCAACATGG 3231

Qy   571 CGAAACCTCATTTCTTACTATAAAAATACAAAAACTAACCAGGCGCTGGTGTGGCTTGTCCTGGTA 630
          ||||| | | || | || ||||||||||||||||| ||||||||| ||||| || |||| ||||
Db   3232 TGAAAACCCCGTYCTCTACTAAAAAATACAAAAAAATTAGCCAGGCATGGTGGCGCATTTGCCTGTA 3291

Qy   631 ATCCCAGCTACTCGGGAGGCTGAGGCAGGAGGAATCGGCTTGAACCGGGAGGCGGAGGTTG 690
          |||||||||||||||||||||||||||||| ||||| ||||||||| |||||||||||||
Db   3292 ATCCCAGCTACTCGGGAGGCTGAGGCAGGAGAGAATCACTTGAACCTGGGAGGCGGAGGTTG 3351

Qy   691 CAGTGAGCTGAGATCACACCACCACTGCACTCCAGCCTGGGCGACAGAGCGAGACTCCAGCTT 750
          |||||||||||||||| | ||||||||| |||| ||||||| |||||| |||||||| |
Db   3352 CAGTGAGCTGAGATCACGCCACTGCACTCTAGCCCTGGGCAACAGAGTGAGACTCCATCTC 3411

Qy   751 AAAAAAAAAAAAAAAAAAAAAAAAAAGAGACCATCACTGCGTCTCCTCCATTCTTACAGAT 810
          ||||| || ||| |
Db   3412 AAAAACTGAAAATGAAAATAAAAATATGTAACTGAAATATTTACTTAATCTG 3471

Qy   811 GAAAAAACAGGCTCAGAGGTTGAATCGTTTTCCTGAAGTC 850
          ||||| | ||| ||| || || | | ||
Db   3472 GAAAACAAATGTAACTAATTTTTAAAGTGGTTACATCTATTC 3511

RESULT 13
ABA17151
ID   ABA17151 standard; DNA; 2981 BP.
XX
AC   ABA17151;
XX
DT   23-JAN-2002 (first entry)
XX

DE   Human nervous system related polynucleotide SEQ ID NO 9482.
XX
KW   Human; nootropic; neuroprotective; cytostatic; dermatological; virucide;
KW   Immunosuppressive; antiinflammatory; anti-HIV; antibacterial; vulnerary;
KW   antiparkinsonian; antisickling; antianaemic; antiarthritic; cancer;
KW   antirheumatic; hepatotropic; cerebroprotective; antiinflammatory;
KW   antiallergic; antidiabetic; antiulcer; anticonvulsant; antifungal;
KW   antiparasitic; cardiant; immune disorder; cardiovascular disorder;
KW   neurological disease; infection; nephrotropic; gene therapy; vaccine; ds.
XX
OS   Homo sapiens.
XX
PN   WO200159063-A2.
XX
PD   16-AUG-2001.
XX
PF   17-JAN-2001; 2001WO-US01334.
XX
PR   31-JAN-2000; 2000US-0179065.
PR   04-FEB-2000; 2000US-0180628.
PR   24-FEB-2000; 2000US-0184664.
PR   02-MAR-2000; 2000US-0186350.
PR   02-MAR-2000; 2000US-0189874.
PR   16-MAR-2000; 2000US-0190076.
PR   17-MAR-2000; 2000US-0198123.
PR   18-APR-2000; 2000US-0205515.
PR   19-MAY-2000; 2000US-0209467.
PR   07-JUN-2000; 2000US-0214886.
PR   07-JUN-2000; 2000US-0215135.
PR   28-JUN-2000; 2000US-0216647.
PR   30-JUN-2000; 2000US-0216880.
PR   07-JUL-2000; 2000US-0217487.
PR   07-JUL-2000; 2000US-0217496.
PR   11-JUL-2000; 2000US-0218290.
PR   11-JUL-2000; 2000US-0220963.
PR   14-JUL-2000; 2000US-0220964.
PR   26-JUL-2000; 2000US-0224518.
PR   26-JUL-2000; 2000US-0224519.
PR   14-AUG-2000; 2000US-0225213.
PR   14-AUG-2000; 2000US-0225214.
PR   14-AUG-2000; 2000US-0225266.
PR   14-AUG-2000; 2000US-0225267.
PR   14-AUG-2000; 2000US-0225268.
PR   14-AUG-2000; 2000US-0225270.
PR   14-AUG-2000; 2000US-0225447.
PR   14-AUG-2000; 2000US-0225757.
PR   14-AUG-2000; 2000US-0225758.
PR   14-AUG-2000; 2000US-0225759.
PR   18-AUG-2000; 2000US-0226279.
PR   22-AUG-2000; 2000US-0226681.
PR   22-AUG-2000; 2000US-0226868.
PR   23-AUG-2000; 2000US-0227009.
PR   30-AUG-2000; 2000US-0227182.
PR   01-SEP-2000; 2000US-0228924.
PR   01-SEP-2000; 2000US-0229287.
PR   01-SEP-2000; 2000US-0229343.
PR   01-SEP-2000; 2000US-0229344.
PR   05-SEP-2000; 2000US-0229345.
PR   05-SEP-2000; 2000US-0229509.
PR   06-SEP-2000; 2000US-0229513.
PR   06-SEP-2000; 2000US-0230437.
PR   06-SEP-2000; 2000US-0230438.
PR   08-SEP-2000; 2000US-0231242.
PR   08-SEP-2000; 2000US-0231243.
PR   08-SEP-2000; 2000US-0231244.
PR   08-SEP-2000; 2000US-0231413.
PR   08-SEP-2000; 2000US-0231414.
PR   08-SEP-2000; 2000US-0232080.
PR   08-SEP-2000; 2000US-0232081.
PR   12-SEP-2000; 2000US-0231968.
PR   14-SEP-2000; 2000US-0232397.
PR   14-SEP-2000; 2000US-0232398.
PR   14-SEP-2000; 2000US-0232399.

PR   14-SEP-2000; 2000US-0232400.
PR   14-SEP-2000; 2000US-0232401.
PR   14-SEP-2000; 2000US-0233063.
PR   14-SEP-2000; 2000US-0233064.
PR   14-SEP-2000; 2000US-0233065.
PR   21-SEP-2000; 2000US-0234223.
PR   21-SEP-2000; 2000US-0234274.
PR   25-SEP-2000; 2000US-0234497.
PR   25-SEP-2000; 2000US-0234998.
PR   26-SEP-2000; 2000US-0235484.
PR   27-SEP-2000; 2000US-0235834.
PR   27-SEP-2000; 2000US-0235836.
PR   29-SEP-2000; 2000US-0236327.
PR   29-SEP-2000; 2000US-0236367.
PR   29-SEP-2000; 2000US-0236368.
PR   29-SEP-2000; 2000US-0236369.
PR   29-SEP-2000; 2000US-0236370.
PR   02-OCT-2000; 2000US-0236802.
PR   02-OCT-2000; 2000US-0237037.
PR   02-OCT-2000; 2000US-0237038.
PR   02-OCT-2000; 2000US-0237039.
PR   02-OCT-2000; 2000US-0237040.
PR   13-OCT-2000; 2000US-0239935.
PR   13-OCT-2000; 2000US-0239937.
PR   20-OCT-2000; 2000US-0239960.
PR   20-OCT-2000; 2000US-0241785.
PR   20-OCT-2000; 2000US-0241786.
PR   20-OCT-2000; 2000US-0241787.
PR   20-OCT-2000; 2000US-0241808.
PR   20-OCT-2000; 2000US-0241809.
PR   20-OCT-2000; 2000US-0241826.
PR   20-OCT-2000; 2000US-0242221.
PR   01-NOV-2000; 2000US-0242222.
PR   01-NOV-2000; 2000US-0244617.
PR   08-NOV-2000; 2000US-0244474.
PR   08-NOV-2000; 2000US-0244475.
PR   08-NOV-2000; 2000US-0244476.
PR   08-NOV-2000; 2000US-0244477.
PR   08-NOV-2000; 2000US-0244478.
PR   08-NOV-2000; 2000US-0245523.
PR   08-NOV-2000; 2000US-0245524.
PR   08-NOV-2000; 2000US-0245525.
PR   08-NOV-2000; 2000US-0245526.
PR   08-NOV-2000; 2000US-0245527.
PR   08-NOV-2000; 2000US-0245528.
PR   08-NOV-2000; 2000US-0246532.
PR   08-NOV-2000; 2000US-0246609.
PR   08-NOV-2000; 2000US-0246610.
PR   08-NOV-2000; 2000US-0246611.
PR   17-NOV-2000; 2000US-0246613.
PR   17-NOV-2000; 2000US-0249207.
PR   17-NOV-2000; 2000US-0249208.
PR   17-NOV-2000; 2000US-0249209.
PR   17-NOV-2000; 2000US-0249210.
PR   17-NOV-2000; 2000US-0249211.
PR   17-NOV-2000; 2000US-0249212.
PR   17-NOV-2000; 2000US-0249213.
PR   17-NOV-2000; 2000US-0249214.
PR   17-NOV-2000; 2000US-0249215.
PR   17-NOV-2000; 2000US-0249216.
PR   17-NOV-2000; 2000US-0249217.
PR   17-NOV-2000; 2000US-0249218.
PR   17-NOV-2000; 2000US-0249244.
PR   17-NOV-2000; 2000US-0249245.
PR   17-NOV-2000; 2000US-0249264.
PR   17-NOV-2000; 2000US-0249265.
PR   17-NOV-2000; 2000US-0249297.
PR   17-NOV-2000; 2000US-0249299.
PR   17-NOV-2000; 2000US-0249300.
PR   01-DEC-2000; 2000US-0250391.
PR   01-DEC-2000; 2000US-0251030.
PR   05-DEC-2000; 2000US-0251160.
PR   05-DEC-2000; 2000US-0251988.
PR   05-DEC-2000; 2000US-0256719.

```
PR   06-DEC-2000; 200US-0251479.
PR   08-DEC-2000; 200US-0251856.
PR   08-DEC-2000; 200US-0251868.
PR   08-DEC-2000; 200US-0251869.
PR   08-DEC-2000; 200US-0251989.
PR   08-DEC-2000; 200US-0251990.
PR   11-DEC-2000; 200US-0254097.
PR   05-JAN-2001; 2001US-0259678.
XX
PA   (HUMA-) HUMAN GENOME SCI INC.
XX
PI   Rosen CA, Barash SC, Ruben SM;
XX
DR   WPI; 2001-541565/60.
XX
PT   Nucleic acids encoding 3224 human nervous system antigen polypeptides,
PT   useful for preventing, diagnosing and/or treating nervous system
PT   cancers and metastases -
XX
PS   Disclosure; SEQ ID NO 9482; 170lpp + Sequence Listing; English.
XX
CC   The invention relates to novel genes (ABA11004-ABA21534) and proteins
CC   (ABB14678-ABB18001) useful for preventing, treating or ameliorating
CC   medical conditions e.g. by protein or gene therapy. The genes are
CC   isolated from a range of human tissues disclosed in the specification.
CC   The nucleic acids, proteins, antibodies and (antagonists are useful
CC   in the diagnosis, treatment and prevention of: (a) cancer, e.g. breast
CC   and ovarian cancer and other cancers of the adrenal gland, bone, bone
CC   marrow, breast, gastrointestinal tract, liver, lung, or urogenital;
CC   (b) immune disorders e.g. Addison's disease, allergies, autoimmune
CC   haemolytic anaemia, autoimmune thyroiditis, diabetes mellitus, Crohn's
CC   disease, multiple sclerosis, rheumatoid arthritis and ulcerative
CC   colitis; (c) cardiovascular disorders such as myocardial ischaemias;
CC   (d) wound healing; (e) neurological diseases e.g. cerebral anoxia and
CC   epilepsy; and (f) infectious diseases such as viral, bacterial, fungal
CC   and parasitic infections.
CC   Note: The sequence data for this patent did not form part of the
CC   printed specification, but was obtained in electronic format directly
CC   from WIPO at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ   Sequence 2981 BP; 841 A; 695 C; 652 G; 793 T; 0 other;

Query Match             24.8%;  Score 247.8;  DB 22;  Length 2981;
Best Local Similarity   85.48%;
Matches  276;  Conservative    0;  Mismatches   47;  Indels    0;  Gaps    0;
Pred. No. 2.4e-51;

Qy     450 CTCTCAAGAGACAGTCATTGGGCTGGGCACGGTGGCCTCATGCCTGCAATCCCAGCACCTTG 509
           ||||| ||   |||||| |||||||| ||||||||||||||||||| |||||||||||| |||
Db     928 CTCTTAAAAGAACCTTTCCAGCCGCGGCCCGCCAGTGGCTCATGCCTGTAATCCCAGCACTTTG 987

Qy     510 GGAGGCTGAGGCAGGCAGATGACTTGAGGCCCAGGAGTTCGAGACCAGCCTGGCCAATATG 569
           |||| |||||||||||||||| |||||||| |||||||||||||||||||||||| || |||
Db     988 GGAGGCCGAGGCAGGCAGATGACTTGAGGTCAGGAGTTCGAGACCAGCCTGGCCAACGTG 1047

Qy     570 GCGGAAACCTCATTTCTACTAAAAATACAAAAAACTAACCAGGCGTGGTGGCTGTGCCTGT 629
           || |||||||| || |||||||||||||||||    ||| |||||| || |||||||||
Db    1048 GCAAAACCCCATCTCTACTAAAAATACAAAAAATTAGCCGGGTGTGGTGGCACATGCCTGT 1107

Qy     630 AATCCCAGCTACTCGGGAGGCTGAGGCAGGAGAATCGCTTGAACCGGGAGGCAGAGGTT 689
           |||||||||||||| ||||||||||||||||||||||||||||||   |||||||| |||
Db    1108 AATCCCAGCTACTCAGGAGGCTGAGGCAGGAGAATCGCTTGAACCCAGGAGGCAGAGATT 1167

Qy     690 GCAGTGAGCTGAGATCACACACCTGCACTCCAGCCTGGGCAGCACAGAGCGAGACTCCAGCT 749
           |||||||||  | ||| ||| ||||||||||||||| || | |||||||  |||||||| ||
Db    1168 GCAGTGAGCCAAGATCGCACCACTGCACTCCAGCCTCGGTGACAGAGCAAGACTCCATCT 1227

Qy     750 TAAAAAAAAAAAAAAAAAAAAAA 772
            | |||||||||||||||||||||
Db    1228 CAGGAAAAAAAAAAAAAAAAAAAA 1250
```

```
ID   AAL05811 standard; DNA; 3172 BP.
XX
AC   AAL05811;
XX
DT   21-NOV-2001 (first entry)
XX
DE   Human reproductive system related antigen DNA SEQ ID NO: 8499.
XX
KW   Human; reproductive system related antigen; reproductive system disorder;
KW   cancer; gene therapy; ds.
XX
OS   Homo sapiens.
XX
PN   WO200155320-A2.
XX
PD   02-AUG-2001.
XX
PF   17-JAN-2001; 2001WO-US01339.
XX
PR   31-JAN-2000; 2000US-0179065.
PR   04-FEB-2000; 2000US-0180628.
PR   24-FEB-2000; 2000US-0184664.
PR   02-MAR-2000; 2000US-0186350.
PR   16-MAR-2000; 2000US-0189874.
PR   17-MAR-2000; 2000US-0190076.
PR   18-APR-2000; 2000US-0199123.
PR   19-MAY-2000; 2000US-0205515.
PR   07-JUN-2000; 2000US-0209467.
PR   28-JUN-2000; 2000US-0214886.
PR   30-JUN-2000; 2000US-0215135.
PR   07-JUL-2000; 2000US-0216647.
PR   07-JUL-2000; 2000US-0216880.
PR   11-JUL-2000; 2000US-0217487.
PR   11-JUL-2000; 2000US-0217496.
PR   14-JUL-2000; 2000US-0218290.
PR   26-JUL-2000; 2000US-0220963.
PR   26-JUL-2000; 2000US-0220964.
PR   14-AUG-2000; 2000US-0224418.
PR   14-AUG-2000; 2000US-0224519.
PR   14-AUG-2000; 2000US-0225213.
PR   14-AUG-2000; 2000US-0225214.
PR   14-AUG-2000; 2000US-0225266.
PR   14-AUG-2000; 2000US-0225267.
PR   14-AUG-2000; 2000US-0225268.
PR   14-AUG-2000; 2000US-0225270.
PR   14-AUG-2000; 2000US-0225447.
PR   14-AUG-2000; 2000US-0225757.
PR   14-AUG-2000; 2000US-0225758.
PR   14-AUG-2000; 2000US-0225759.
PR   18-AUG-2000; 2000US-0226279.
PR   22-AUG-2000; 2000US-0226681.
PR   22-AUG-2000; 2000US-0226868.
PR   22-AUG-2000; 2000US-0227182.
PR   23-AUG-2000; 2000US-0227009.
PR   30-AUG-2000; 2000US-0228924.
PR   01-SEP-2000; 2000US-0229287.
PR   01-SEP-2000; 2000US-0229333.
PR   01-SEP-2000; 2000US-0229344.
PR   01-SEP-2000; 2000US-0229345.
PR   05-SEP-2000; 2000US-0229509.
PR   05-SEP-2000; 2000US-0229513.
PR   06-SEP-2000; 2000US-0230437.
PR   06-SEP-2000; 2000US-0230438.
PR   08-SEP-2000; 2000US-0231242.
PR   08-SEP-2000; 2000US-0231243.
PR   08-SEP-2000; 2000US-0231244.
PR   08-SEP-2000; 2000US-0231413.
PR   08-SEP-2000; 2000US-0231414.
PR   08-SEP-2000; 2000US-0232080.
PR   08-SEP-2000; 2000US-0232081.
PR   12-SEP-2000; 2000US-0231968.
PR   14-SEP-2000; 2000US-0232397.
PR   14-SEP-2000; 2000US-0232398.
```

RESULT 14
AAL05811

```
PR   14-SEP-2000;  2000US-0232399.
PR   14-SEP-2000;  2000US-0232400.
PR   14-SEP-2000;  2000US-0232401.
PR   14-SEP-2000;  2000US-0233063.
PR   14-SEP-2000;  2000US-0233064.
PR   14-SEP-2000;  2000US-0233065.
PR   21-SEP-2000;  2000US-0234223.
PR   21-SEP-2000;  2000US-0234274.
PR   21-SEP-2000;  2000US-0234997.
PR   25-SEP-2000;  2000US-0234998.
PR   26-SEP-2000;  2000US-0235484.
PR   27-SEP-2000;  2000US-0235834.
PR   27-SEP-2000;  2000US-0235836.
PR   29-SEP-2000;  2000US-0236327.
PR   29-SEP-2000;  2000US-0236367.
PR   29-SEP-2000;  2000US-0236368.
PR   29-SEP-2000;  2000US-0236369.
PR   29-SEP-2000;  2000US-0236370.
PR   02-OCT-2000;  2000US-0236802.
PR   02-OCT-2000;  2000US-0237037.
PR   02-OCT-2000;  2000US-0237038.
PR   02-OCT-2000;  2000US-0237039.
PR   02-OCT-2000;  2000US-0237040.
PR   13-OCT-2000;  2000US-0239935.
PR   13-OCT-2000;  2000US-0239937.
PR   20-OCT-2000;  2000US-0240960.
PR   20-OCT-2000;  2000US-0241221.
PR   20-OCT-2000;  2000US-0241785.
PR   20-OCT-2000;  2000US-0241786.
PR   20-OCT-2000;  2000US-0241787.
PR   20-OCT-2000;  2000US-0241808.
PR   20-OCT-2000;  2000US-0241809.
PR   20-OCT-2000;  2000US-0241826.
PR   01-NOV-2000;  2000US-0244617.
PR   08-NOV-2000;  2000US-0246474.
PR   08-NOV-2000;  2000US-0246475.
PR   08-NOV-2000;  2000US-0246476.
PR   08-NOV-2000;  2000US-0246477.
PR   08-NOV-2000;  2000US-0246478.
PR   08-NOV-2000;  2000US-0246523.
PR   08-NOV-2000;  2000US-0246524.
PR   08-NOV-2000;  2000US-0246525.
PR   08-NOV-2000;  2000US-0246526.
PR   08-NOV-2000;  2000US-0246527.
PR   08-NOV-2000;  2000US-0246528.
PR   08-NOV-2000;  2000US-0246532.
PR   08-NOV-2000;  2000US-0246609.
PR   08-NOV-2000;  2000US-0246610.
PR   08-NOV-2000;  2000US-0246611.
PR   08-NOV-2000;  2000US-0246613.
PR   17-NOV-2000;  2000US-0249207.
PR   17-NOV-2000;  2000US-0249208.
PR   17-NOV-2000;  2000US-0249209.
PR   17-NOV-2000;  2000US-0249210.
PR   17-NOV-2000;  2000US-0249211.
PR   17-NOV-2000;  2000US-0249212.
PR   17-NOV-2000;  2000US-0249213.
PR   17-NOV-2000;  2000US-0249214.
PR   17-NOV-2000;  2000US-0249215.
PR   17-NOV-2000;  2000US-0249216.
PR   17-NOV-2000;  2000US-0249217.
PR   17-NOV-2000;  2000US-0249218.
PR   17-NOV-2000;  2000US-0249244.
PR   17-NOV-2000;  2000US-0249264.
PR   17-NOV-2000;  2000US-0249265.
PR   17-NOV-2000;  2000US-0249299.
PR   17-NOV-2000;  2000US-0249300.
PR   01-DEC-2000;  2000US-0250160.
PR   01-DEC-2000;  2000US-0250391.
PR   05-DEC-2000;  2000US-0251030.
PR   05-DEC-2000;  2000US-0251988.

PR   05-DEC-2000;  2000US-0256719.
PR   06-DEC-2000;  2000US-0251479.
PR   08-DEC-2000;  2000US-0251856.
PR   08-DEC-2000;  2000US-0251868.
PR   08-DEC-2000;  2000US-0251869.
PR   08-DEC-2000;  2000US-0251989.
PR   08-DEC-2000;  2000US-0251990.
PR   11-DEC-2000;  2000US-0254097.
PR   05-JAN-2001;  2001US-0259678.
XX
PA   (HUMA-) HUMAN GENOME SCI INC.
XX
PI   Rosen CA, Barash SC, Ruben SM;
XX
DR   WPI; 2001-465570/50.
XX
PT   Isolated nucleic acid molecule encoding a reproductive system antigen
PT   is used in preventing, treating or ameliorating a medical condition -
XX
PS   Disclosure; SEQ ID NO 8499; 1297pp + Sequence Listing; English.
XX
CC   The present invention provides the protein and coding sequences of a
CC   number of human reproductive system related antigens. These can be used
CC   in the prevention and treatment of reproductive system disorders,
CC   including cancer. The present sequence is a genomic sequence encoding a
CC   protein of the invention.
XX
SQ   Sequence 3172 BP; 1024 A; 529 C; 711 G; 908 T; 0 other;
XX

Query Match             24.8%;  Score 247.8;  DB 22;  Length 3172;
Best Local Similarity   85.4%;  Pred. No. 2.4e-51;
Matches  276;  Conservative   0;  Mismatches  47;  Indels   0;  Gaps   0;

Qy    455 AAGAGACAGTCATTGGCTGGGCACGGTGGCTCATGCCTGCAATCCCAGCACCTTGGGAGG 514
          ||||||| | ||||| |||||| |||||||||||||||||||||||||||||||||||||
Db   2219 AAAAAAAAAAGAGAGAGGCTGGGGTCGGGGTGGGCTCACGCCGCCTTATAATCCCAGCAGG 2278

Qy    515 CTGAGGCAGGCAGATGACTTGAGCCCAGGAGTTCGAGACCAGCCTGGCCAATATGGCGAA 574
          | |||||||||| | ||||| ||||||||||||||||| | ||||||||||||||||| |
Db   2279 CAGAGGCAGGCAGATCACCTGAGGTCAGGAGTTCAAGACCAGCCTGGGCAACATGGTGAA 2338

Qy    575 ACCTCATTTCTACTAAAAATACAAAAACTAACCAGGCGTGGTGGCTTGTGCCTGTAATCC 634
          | |||| |||||||||||||||||||| | |||||||||||||||| | | |||||||||
Db   2339 ACCCCATCTCTACTAAAAATACAAAAATTAGCTGGGCGTGGTGGGTACCTGTAATCC 2398

Qy    635 CAGCTACTCGGGAGGCTGAGGCAGGAGAATCGCTTGAACCGGGGAGGCGGAGGTTGCAGT 694
          |||| || ||||||||||||||||||||| |||||||||| | | |||| ||||||||||
Db   2399 CAGCTCCTTGGGAGGCTGAGGCAGGAGAATCTCTTGAACCTGGGAGGAAGAGATTGCAGT 2458

Qy    695 GAGCTGAGATCACACCACTGCACTCCAGCCTGGGCGACAGAGCGAGACTCCAGCTTAAA 754
          |||||||||||| | |||||| | |||||| ||||| |||||| |||||||| || |||
Db   2459 GAGCTGAGATCATGCCACTGCTACTCCAGCACTGGGTGACAGAGTGAGACTCCATCTCAAAA 2518

Qy    755 AAAAAAAAAAAAAAAAAGG 777
          |||||||||||||||||
Db   2519 AAAAAAAAAAAAAAAAATG 2541
```

RESULT 15
AAK85276/c
```
ID   AAK85276 standard; DNA; 3172 BP.
XX
AC   AAK85276;
XX
DT   07-NOV-2001 (first entry)
XX
DE   Human immune/haematopoietic antigen genomic sequence SEQ ID NO:40088.
XX
KW   Human; immune; haematopoietic; immune/haematopoietic antigen; cancer;
KW   cytostatic; gene therapy; vaccine; metastasis; ds.
XX
OS   Homo sapiens.
XX
```

```
PN   WO20015182-A2.
XX
PD   09-AUG-2001.
XX
PF   17-JAN-2001; 2001WO-US01354.
XX
PR   31-JAN-2000; 2000US-0179065.
PR   04-FEB-2000; 2000US-0180628.
PR   24-FEB-2000; 2000US-0184664.
PR   02-MAR-2000; 2000US-0186350.
PR   16-MAR-2000; 2000US-0189874.
PR   17-MAR-2000; 2000US-0190076.
PR   18-APR-2000; 2000US-0198123.
PR   19-MAY-2000; 2000US-0205515.
PR   07-JUN-2000; 2000US-0209467.
PR   28-JUN-2000; 2000US-0214886.
PR   30-JUN-2000; 2000US-0215135.
PR   07-JUL-2000; 2000US-0216647.
PR   07-JUL-2000; 2000US-0216880.
PR   11-JUL-2000; 2000US-0217487.
PR   11-JUL-2000; 2000US-0217496.
PR   14-JUL-2000; 2000US-0218290.
PR   26-JUL-2000; 2000US-0220963.
PR   26-JUL-2000; 2000US-0220964.
PR   14-AUG-2000; 2000US-0224518.
PR   14-AUG-2000; 2000US-0224519.
PR   14-AUG-2000; 2000US-0225213.
PR   14-AUG-2000; 2000US-0225214.
PR   14-AUG-2000; 2000US-0225266.
PR   14-AUG-2000; 2000US-0225267.
PR   14-AUG-2000; 2000US-0225268.
PR   14-AUG-2000; 2000US-0225270.
PR   14-AUG-2000; 2000US-0225447.
PR   14-AUG-2000; 2000US-0225757.
PR   14-AUG-2000; 2000US-0225758.
PR   14-AUG-2000; 2000US-0225759.
PR   18-AUG-2000; 2000US-0226279.
PR   18-AUG-2000; 2000US-0226681.
PR   22-AUG-2000; 2000US-0226868.
PR   22-AUG-2000; 2000US-0227182.
PR   23-AUG-2000; 2000US-0227009.
PR   30-AUG-2000; 2000US-0228924.
PR   01-SEP-2000; 2000US-0229287.
PR   01-SEP-2000; 2000US-0229343.
PR   01-SEP-2000; 2000US-0229344.
PR   01-SEP-2000; 2000US-0229345.
PR   05-SEP-2000; 2000US-0229509.
PR   05-SEP-2000; 2000US-0229513.
PR   06-SEP-2000; 2000US-0230437.
PR   06-SEP-2000; 2000US-0230438.
PR   08-SEP-2000; 2000US-0231242.
PR   08-SEP-2000; 2000US-0231243.
PR   08-SEP-2000; 2000US-0231244.
PR   08-SEP-2000; 2000US-0231413.
PR   08-SEP-2000; 2000US-0231414.
PR   08-SEP-2000; 2000US-0232080.
PR   08-SEP-2000; 2000US-0232081.
PR   12-SEP-2000; 2000US-0231968.
PR   14-SEP-2000; 2000US-0232397.
PR   14-SEP-2000; 2000US-0232398.
PR   14-SEP-2000; 2000US-0232399.
PR   14-SEP-2000; 2000US-0232400.
PR   14-SEP-2000; 2000US-0232401.
PR   14-SEP-2000; 2000US-0233063.
PR   14-SEP-2000; 2000US-0233064.
PR   14-SEP-2000; 2000US-0233065.
PR   21-SEP-2000; 2000US-0234223.
PR   21-SEP-2000; 2000US-0234274.
PR   25-SEP-2000; 2000US-0234997.
PR   25-SEP-2000; 2000US-0234998.
PR   26-SEP-2000; 2000US-0235484.
PR   27-SEP-2000; 2000US-0235834.
PR   27-SEP-2000; 2000US-0235836.

PR   29-SEP-2000; 2000US-0236227.
PR   29-SEP-2000; 2000US-0236367.
PR   29-SEP-2000; 2000US-0236368.
PR   29-SEP-2000; 2000US-0236369.
PR   29-SEP-2000; 2000US-0236370.
PR   02-OCT-2000; 2000US-0236802.
PR   02-OCT-2000; 2000US-0237037.
PR   02-OCT-2000; 2000US-0237038.
PR   02-OCT-2000; 2000US-0237039.
PR   02-OCT-2000; 2000US-0237040.
PR   13-OCT-2000; 2000US-0239935.
PR   13-OCT-2000; 2000US-0239937.
PR   20-OCT-2000; 2000US-0240960.
PR   20-OCT-2000; 2000US-0241221.
PR   20-OCT-2000; 2000US-0241785.
PR   20-OCT-2000; 2000US-0241786.
PR   20-OCT-2000; 2000US-0241787.
PR   20-OCT-2000; 2000US-0241808.
PR   20-OCT-2000; 2000US-0241809.
PR   20-OCT-2000; 2000US-0241826.
PR   01-NOV-2000; 2000US-0244617.
PR   08-NOV-2000; 2000US-0246474.
PR   08-NOV-2000; 2000US-0246475.
PR   08-NOV-2000; 2000US-0246476.
PR   08-NOV-2000; 2000US-0246477.
PR   08-NOV-2000; 2000US-0246478.
PR   08-NOV-2000; 2000US-0246523.
PR   08-NOV-2000; 2000US-0246524.
PR   08-NOV-2000; 2000US-0246525.
PR   08-NOV-2000; 2000US-0246526.
PR   08-NOV-2000; 2000US-0246527.
PR   08-NOV-2000; 2000US-0246528.
PR   08-NOV-2000; 2000US-0246532.
PR   08-NOV-2000; 2000US-0246609.
PR   08-NOV-2000; 2000US-0246610.
PR   08-NOV-2000; 2000US-0246611.
PR   08-NOV-2000; 2000US-0246613.
PR   17-NOV-2000; 2000US-0249207.
PR   17-NOV-2000; 2000US-0249208.
PR   17-NOV-2000; 2000US-0249209.
PR   17-NOV-2000; 2000US-0249210.
PR   17-NOV-2000; 2000US-0249211.
PR   17-NOV-2000; 2000US-0249212.
PR   17-NOV-2000; 2000US-0249213.
PR   17-NOV-2000; 2000US-0249214.
PR   17-NOV-2000; 2000US-0249215.
PR   17-NOV-2000; 2000US-0249216.
PR   17-NOV-2000; 2000US-0249217.
PR   17-NOV-2000; 2000US-0249218.
PR   17-NOV-2000; 2000US-0249244.
PR   17-NOV-2000; 2000US-0249245.
PR   17-NOV-2000; 2000US-0249264.
PR   17-NOV-2000; 2000US-0249265.
PR   17-NOV-2000; 2000US-0249297.
PR   17-NOV-2000; 2000US-0249299.
PR   17-NOV-2000; 2000US-0249300.
PR   01-DEC-2000; 2000US-0250160.
PR   01-DEC-2000; 2000US-0250391.
PR   05-DEC-2000; 2000US-0251030.
PR   05-DEC-2000; 2000US-0251479.
PR   05-DEC-2000; 2000US-0251719.
PR   06-DEC-2000; 2000US-0251988.
PR   08-DEC-2000; 2000US-0251856.
PR   08-DEC-2000; 2000US-0251868.
PR   08-DEC-2000; 2000US-0251869.
PR   08-DEC-2000; 2000US-0251989.
PR   08-DEC-2000; 2000US-0251990.
PR   11-DEC-2000; 2000US-0254097.
PR   05-JAN-2001; 2001US-0259678.
XX
PA   (HUMA-) HUMAN GENOME SCI INC.
XX
PI   Rosen CA,  Barash SC,  Ruben SM;
```

```
XX
DR    WPI;  2001-483426/52.
XX
PT    Nucleic acids encoding human immune/hematopoietic antigen polypeptides,
PT    useful for preventing, diagnosing and/or treating cancers and
PT    metastasis -
XX
PS    Disclosure;  SEQ ID NO 40088;  307lpp + Sequence Listing; English.
XX
CC
CC    AAK54951 to AAK64702 encode the human immune/haematopoietic antigen (I)
CC    amino acid sequences given in AAM82170 to AAM91921. (I) have cytostatic
CC    activity, and can be used in gene therapy and vaccine production. (I)
CC    proteins and polynucleotides may be used in the prevention, diagnosis and
CC    treatment of diseases associated with inappropriate (I) expression. For
CC    example, they may be used to treat disorders associated with decreased
CC    expression by rectifying mutations or deletions in a patient's genome
CC    that affect the activity of (I) by expressing inactive proteins or to
CC    supplement the patients own production of (I). Additionally, (I)
CC    polynucleotides may be used to produce the secreted (I), by inserting
CC    the nucleic acids into a host cell and culturing the cell to express the
CC    protein. (I) proteins and polynucleotides may be used to prevent,
CC    diagnose and treat immune/haematopoietic-related diseases, especially
CC    cancers and cancer metastases of haematopoietic-derived cells. AAK64703
CC    to AAK87694 represent human immune/haematopoietic antigen genomic
CC    sequences from the present invention. AAK54942 to AAK54950 and AAM82169
CC    represent sequences used in the exemplification of the present invention.
XX
SQ    Sequence 3172 BP; 908 A; 711 C; 529 G; 1024 T; 0 other;

Query Match             24.8%;  Score 247.8;  DB 22;  Length 3172;
Best Local Similarity   85.4%;  Pred. No. 2.4e-51;
Matches  276;  Conservative    0;  Mismatches   47;  Indels    0;  Gaps   0;

Qy     455 AAGAGACAGTCATTGGCTGGGCCACGGTGGCCTCATGGCTCAAATCCCAGCACCTTTGGAGG 514
           || ||| |  | ||||| |||| ||| | ||||||||||||  ||||||||| || |||||
Db     954 AAAAAAAAAGAGAGGCTGGGTGGGTTGCCGGTGGCCTCACGGCCTATAATCCCACACTTTGGGGAGG 895

Qy     515 CTGAGGCAGGCAGATGACTTGAGGCCAGGAGTTCGAGACCAGCCCTGCCCAATATGGCGAA 574
           |  ||| ||||||||  |||||| ||||||||||  |||||||||| || |||||   ||
Db     894 CAGAGGCAGGCAGATCACCTGAGGTCAGGAGTTCAAGACCACCCCTGCCCACACATGGTGAA 835

Qy     575 ACCTCATTTCTACTAAAAATACAAAAAAACTAACCAGGCGTGGTGGCTTGTGCCTGTAATCC 634
           ||  ||  |||||||| | ||||  |   || |  | |||||  | ||| | |||||||||
Db     834 ACCCCATCTCTACTAAAAATACAAAAAATTAGCTGGGCGTGGTGGGGTACCTGTAATCC 775

Qy     635 CAGCTACTCGGGAGGCTGAGGCAGGAGAATGCCTTGAACCCGGGAGGCGAGGTTGCAGT 694
           ||||  ||  |||||||||||| || |||||||||||||||| ||||||||||||| |||
Db     774 CAGCTCCTTGGGAGGCTGAGGCAGGAGAATCTCTTGAACCTGGGAGGAAGAGAATTGCAGT 715

Qy     695 GAGCTGAGATCACACCACTGCACTCCAGCCTGGGCGACAGAGCGAGACTCCAGCTTAAAA 754
           |||| |||||| |||| || |  || |||| ||| || |||| ||||||||||||| ||||
Db     714 GAGCTGAGATCATGCCACTGTPACTCCAGACTGAGTGACAGAGTGAGACTCCATCTCAAAA 655

Qy     755 AAAAAAAAAAAAAAAAAAAAAGG 777
           ||||||||||||||||||||||| |
Db     654 AAAAAAAAAAAAAAAAAAAAATG 632
```

Search completed: March 29, 2003, 17:41:04
Job time : 780.075 secs

GenCore version 5.1.4_p5_4578

Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on:　　　March 29, 2003, 16:20:14 ; Search time 1836 Seconds
　　　　　　　　　　　　　　　　　　　　　(without alignments)
　　　　　　　　　　　　　　　　　　　　　8829.908 Million cell updates/sec

Title:　　　　　US-09-691-219-3_COPY_10000_11000
Perfect score: 1001
Sequence:　　　1 ttactttaccatcctgggg..........ggcccatgggtgaggacatg 1001

Scoring table: IDENTITY_NUC
　　　　　　　　　Gapop 10.0 , Gapext 1.0

Searched:　　　16154066 seqs, 8097743376 residues

Total number of hits satisfying chosen parameters:　　32308132

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
　　　　　　　　　Maximum Match 100%
　　　　　　　　　Listing first 45 summaries

Database :
　　　　1: em_estba:*
　　　　2: em_esthum:*
　　　　3: em_estin:*
　　　　4: em_estmu:*
　　　　5: em_estov:*
　　　　6: em_estpl:*
　　　　7: em_estro:*
　　　　8: em_htc:*
　　　　9: em_est1:*
　　　10: gb_est1:*
　　　11: gb_htc:*
　　　12: gb_est2:*
　　　13: gb_est4:*
　　　14: gb_est5:*
　　　15: gb_estun:*
　　　16: em_estom:*
　　　17: gb_gss:*
　　　18: em_gss_hum:*
　　　19: em_gss_inv:*
　　　20: em_gss_pln:*
　　　21: em_gss_vrt:*
　　　22: em_gss_fun:*
　　　23: em_gss_mam:*
　　　24: em_gss_mus:*
　　　25: em_gss_other:*
　　　26: em_gss_pro:*
　　　27: em_gss_rod:*

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	326.4	32.6	401	9	AU156173	AU156173 AU156173
2	252.8	25.3	624	12	BF854308	BF854308 MR2-EN009
3	251.8	25.2	664	17	AG151502	AG151502 Pan trogl
4	250.2	25.0	642	17	B59854	B59854 CIT-HSP-345
5	248	24.8	1100	14	BM802793	BM802793 AGENCOURT
6	247.6	24.7	711	17	AQ415030	AQ415030 RPCI-11-2

c	7	247.4	24.7	439	9	AA515728	AA515728 ng70f04.s
c	8	247.2	24.7	460	17	B15692	B15692 345N8.TP CI
c	9	246.8	24.7	430	9	AA832145	AA832145 0e22g09.s
c	10	246.6	24.7	416	17	BH367159	BH367159 UP_503-13
c	11	246.6	24.6	782	12	BG743198	BG743198 602634360
c	12	246.4	24.6	530	9	AL704074	AL704074 DKFZp6860
c	13	246	24.6	657	17	B65134	B65134 CIT-HSP-201
c	14	245.8	24.6	521	10	AW970571	AW970571 EST382652
c	15	245.6	24.5	552	14	BM993322	BM993322 UI-H-DT0-
c	16	245.6	24.5	590	17	AQ588588	AQ588588 CITBI-E1-
c	17	245.6	24.5	1273	11	BC011712	BC011712 Homo sapi
c	18	245.4	24.5	910	12	BF337320	BF337320 602034695
c	19	244.8	24.5	437	9	AA644090	AA644090 ab62b04.s
c	20	244.4	24.4	760	9	AL691744	AL691744 DKFZp313M
c	21	244.2	24.4	485	9	AI523316	AI523316 ar71fil.x
c	22	244	24.4	664	17	AQ742068	AQ742068 HS_5566_B
c	23	243.6	24.3	388	10	AW069227	AW069227 cr41h09.x
c	24	243.6	24.3	405	10	AW189113	AW189113 x101b12.x
c	25	243.6	24.3	453	12	BF725844	BF725844 bx20f10.x
c	26	243.6	24.3	504	17	BH152879	BH152879 UP_297-17
c	27	243.6	24.3	620	14	BQ775487	BQ775487 UI-H-FH0-
c	28	243.4	24.3	380	9	AI962030	AI962030 wt42b01.x
c	29	243.4	24.3	551	15	AV756491	AV756491 AV756490
c	30	243.2	24.3	642	17	AQ541546	AQ541546 RPCI-11-3
c	31	243.2	24.3	736	17	AG058933	AG058933 Pan trogl
c	32	243	24.3	355	10	AW084445	AW084445 c58f06.x
c	33	243	24.3	569	17	AQ490878	AQ490878 RPCI-11-2
c	34	242.6	24.3	286	12	BG000416	BG000416 MR3-GN015
c	35	242.6	24.2	286	12	BG000043	BG000043 UI-H-DI0-
c	36	242.6	24.2	706	14	BM989586	BM989586 UI-H-DH0-
c	37	242.6	24.2	844	14	BO961856	BO961856 AGENCOURT
c	38	242.4	24.2	719	17	AQ475181	AQ475181 CITBI-E1-
c	39	242.2	24.2	588	12	BG910226	BG910226 602805702
c	40	242	24.2	575	17	AQ587429	AQ587429 CITBI-E1-
c	41	241.8	24.2	526	17	BE069140	BE069140 QV3-BT037
c	42	241.8	24.2	627	13	BI062936	BI062936 IL3-UT011
c	43	241.8	24.2	678	17	AQ387027	AQ387027 RPCI11-15
c	44	241.6	24.1	634	17	AQ065330	AQ065330 Pan trogl
c	45	241.6	24.1	736	12	BG910839	BG910839 602808596

ALIGNMENTS

RESULT 1
AU156173/c
LOCUS　　　　　AU156173　　　　　　　　　　　401 bp　　　mRNA　　　linear　　EST 05-AUG-2002
DEFINITION　　AU156173 PLACE1 Homo sapiens cDNA clone PLACE1002500 3'', mRNA
　　　　　　　　sequence.
ACCESSION　　　AU156173
VERSION　　　　AU156173.1　　GI:11017694
KEYWORDS　　　　EST.
SOURCE　　　　　human.
ORGANISM　　　　Homo sapiens
　　　　　　　　Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
　　　　　　　　Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE　　　1 (bases 1 to 401)
AUTHORS　　　　Ota,T., Nishikawa,T., Suzuki,Y., Ishii,S., Saito,K., Kawai,Y.,
　　　　　　　　Yamamoto,J., Wakamatsu,A., Ozawa,M., Nakamura,Y., Nagai,T., Sugano
　　　　　　　　,S. and Isogai,T.)
TITLE　　　　　HRI human cDNA project (Ota,T., Nishikawa,T., Suzuki,Y., Ishii,S.,
　　　　　　　　Saito,K., Kawai,Y., Yamamoto,J., Wakamatsu,A., Ozawa,M., Nakamura
　　　　　　　　,Y., Nagai,T., Sugano,S., Isogai,T.)
JOURNAL　　　　Unpublished (2000)
COMMENT　　　　Contact: Takao Isogai
　　　　　　　　Genomics Laboratory
　　　　　　　　Helix Research Institute
　　　　　　　　1532-3 Yana, Kisarazu, Chiba 292-0812, Japan
　　　　　　　　Tel: 81-438-52-3975
　　　　　　　　Fax: 81-438-52-3986
　　　　　　　　Email: genomics@hri.co.jp
　　　　　　　　HRI human cDNA project; 5'- & 3'-end one pass sequencing; Helix
　　　　　　　　Research Institute; cDNA library construction: Department of

Virology, Institute of Medical Science, University of Tokyo, and
Helix Research Institute.

```
FEATURES             Location/Qualifiers
     source          1..401
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="PLACE1002500"
                     /clone_lib="PLACE1"
                     /tissue_type="placenta"
                     /note="Vector: pME18SFL3"
BASE COUNT        87 a      109 c      100 g       96 t       9 others
ORIGIN
```

Query Match   Best Local Similarity 32.6%; Score 326.4; DB 9; Length 401;
Matches 358; Conservative 93.2%; Pred. No. 9.5e-37;
Matches 358; Conservative 0; Mismatches 24; Indels 2; Gaps 2;

```
Qy  368 CAACTACCCCACCCCACCCCAAGGCCATGGCAGGCCCCCAGGAGAACCTTGATCCTGGGCTTTGC 427
        || ||| ||||| || |||||||||||||||||||||||||||||| ||||||| ||||| |||||
Db  383 CANTTTCCCAACCCNACCCCNACCCAAGCCCATGGCAGGCCCCCAGGAATTTGATCTTGGGTTTTGC 324

Qy  428 CGTATGCCAAGTNCCTTACACCCCTCTNCAAGAGAACAGTGAGTCATTGGCTGGGCACGGTGGCTCA 487
        |||||||||||| |||||||| ||| ||||||||||||||||| ||||||||||||||||||||| ||
Db  323 NGTATGCCAAGTCNTTACACCCTTNNCAAGAGAACAGTCATTGGCTGGGCACGGTGGTTCA 264

Qy  488 TGCCTGCCAATCCCAGCACACCTTGGGAGGCCGAGGCGGGCAGGATGACTTGAGGCCCAGGAGTT 547
        |||| || |||||||||| |||||||||||||||||| |||||||||||||||||||||||||||
Db  263 TGCNTGCCAATCCCAGCACCTTGGGAGGCCGAGGCAGGCAGGATGACTTGAGGCCCAGGAGTT 204

Qy  548 CGAGACCAGCCTGGCCAATATGGCGAAAACTCATTTTCTACTAAAAATACAAAAAACTAAACC 607
        ||||||||||||||||||||||||||| |||||||||| |||||||||||||||||||||| ||
Db  203 CGAGACCAGCCTGGCCAATATGGCGAAACNTCATTTTNTACTAAAAATACAAAAAACTAGCC 144

Qy  608 AGGGCTGGGTGGCTTGTGCCTGTAAAT-CCCAGCTACTCGGGAGGCTGAGGCAGGAGAATCG 666
        |||||||| |||||||||||||||||| |||||||||||||||||||||||||||||| |||
Db  143 AGGGCTGGTGGCTTGTGCCTGTAAATCCCCAGCTACTCGGGAGGCTGAGGCAGGAGNTTCG 84

Qy  667 CTTGAACCGGGGAGGCGGAGGTTGCAGTGAGCCGAGATCACACCACTGCACTCCAGCCTG 726
        |||||||||||||||||||||||||||||||||||||||||||| |||||||||||||| ||
Db  83  CTTGAACCGGGGAGGCGGAGGTTGCAGTGAGCCGAGATCACACCACTTGCACTCCAGCCTTG 24

Qy  727 GGCGACGAGACGGAGACTCCAGCTT 750
        || |||| |||| ||||||||||||
Db  23  GG-GACAGAGAAGATTCCAGCTT 1
```

RESULT 2

```
LOCUS       BF854308      624 bp    mRNA    linear    EST 16-JAN-2001
DEFINITION  MR2-EN0092-261200-003-f06 EN0092 Homo sapiens cDNA, mRNA sequence.
ACCESSION   BF854308
VERSION     BF854308.1  GI:12242052
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 624)
  AUTHORS   Dias Neto,E., Garcia Correa,R., Verjovski-Almeida,S., Briones,M.R.,
            Nagai,M.A., da Silva,W.Jr., Zago,M.A., Bordin,S., Costa,F.F.,
            Goldman,G.H., Carvalho,A.F., Matsukuma,A., Baia,G.S., Simpson,D.H.,
            Brunstein,A., deOliveira,P.S., Bucher,P., Jongeneel,C.V., O'Hare
            ,M.J., Soares,F., Brentani,R.R., Reis,L.F., de Souza,S.J. and
            Simpson,A.J.
  TITLE     Shotgun sequencing of the human transcriptome with ORF expressed
            sequence tags
  JOURNAL   Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)
  MEDLINE   20202663
COMMENT     Contact: Simpson A.J.G.
            Ludwig Institute for Cancer Research
            Laboratory of Cancer Genetics
            Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,
            Brazil
            Tel: +55-11-2704922
            Fax: +55-11-2707001
            Email: asimpson@ludwig.org.br
            This sequence was derived from the FAPESP/LICR Human Cancer Genome
            Project. This entry can be seen in the following URL
            (http://www.ludwig.org.br/scripts/gethtml2.pl?ti-MR26t2=MR2-EN0092-
            261200-003-f06&t3=2000-12-26&t4=1)
            Seq primer: puc 18 forward
            High quality sequence start: 8
            High quality sequence stop: 623.
FEATURES             Location/Qualifiers
     source          1..624
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone_lib="EN0092"
                     /note="Organ: lung_normal; Vector: puc18; Site_1: SmaI;
                     Site_2: SmaI; A mini-library was made by cloning products
                     derived from ORESTES PCR (U.S. Letters Patent application
                     No. 196,716 - Ludwig Institute for Cancer Research)
                     profiles into the pUC 18 vector. Reverse transcription of
                     tissue mRNA and cDNA amplification were performed under
                     low stringency conditions."
BASE COUNT       192 a      159 c      163 g      110 t
ORIGIN
```

Query Match   Best Local Similarity 25.3%; Score 252.8; DB 12; Length 624;
Matches 284; Conservative 84.5%; Pred. No. 1.4e-26;
Matches 284; Conservative 0; Mismatches 52; Indels 0; Gaps 0;

```
Qy  455 AAGAGACAGTCATTGGCTGGGCACGGTGGCTCATGCCTGCAATCCCAGCCACCTTGGGAGG 514
        |||| ||||| || ||  | ||||| ||||||||| |||||| |||||||| || ||||||||
Db  201 AAGAAACAGCCAAGGACCAGAGCGCAGTGGCGTCACGCCTGTAATCCCAGCACATTTTGGGAGG 260

Qy  515 CTGAGGCAGGCAGATGACTTGAGGCCAGGAGTTCGAGACCAGCCTGGCCAATATGGCGAA 574
        | ||||||||| |||| |||||||| ||||||||||||||||||||||||||| ||| ||
Db  261 CCGAGGCAGGTGGATCACTTGAGGTCAGGAGTTCGAGACCAGCCTGGCCAACATGGTGAA 320

Qy  575 ACCTCATTTCTACTAAAAATACAAAAAACTACCAACGCGTGGTGGGGGGCGCCTGTAATCC 634
        ||| | | | |||||||||||||| || | |||| || || ||||||||||||||||||||
Db  321 ACCCTGTCTCTACTAAAAATACAATAAGTTAGCCAGGCATGGTGGCAGGCGCCTGTAATCC 380

Qy  635 CAGCTACTCGGGAGGCTGAGGCAGGAGAATCGCTTGAACCGGGAGGCGGAGGTTGCAGT 694
        |||||||| |||||||||||||||||||||| ||||||||| ||||||||||||||||||
Db  381 CAGCTACTCAGGAGGCTGAGGCAGGAGAATTGCCTTGAACCCGGAGGCGGAGGTTGCAGT 440

Qy  695 GAGCTGAGATCACACCACTGCACTCCAGCCTGGGCGACAGAGCGAGACTCCAGCTTAAAA 754
        || | ||||||||||||||||||||||||||||||||| |||||| |||| |||||| |||
Db  441 GACCCAAGATCACACCACTGCACTCCAGCCTGGGCGACAGAGCGAAGACTCATCTCAAAA 500

Qy  755 AAAAAAAAAAAAAAAAAAAAAAAAAAAGGAGAGACCCATCACTGC 790
        ||| |||| || |  ||||  | |||| ||||||||   || |||
Db  501 AAAGAAAAAAGAAAGAAAAAAGAAAAAAAGAAGAAGAAACAGC 536
```

RESULT 3

```
LOCUS       AG151502      664 bp    DNA    linear    GSS 09-JAN-2002
DEFINITION  Pan troglodytes DNA, clone: RP43-014K03.TJ, genomic survey
            sequence.
ACCESSION   AG151502
VERSION     AG151502.1  GI:16681180
KEYWORDS    GSS.
SOURCE      Pan troglodytes
  ORGANISM  Pan troglodytes
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Pan.
REFERENCE   1
  AUTHORS   Fujiyama,A., Hattori,M., Toyoda,A., Taylor,T.D., Yada,T.,
            Totoki,Y., Watanabe,H. and Sakaki,Y.
  TITLE     BAC end sequences of Library RPCI-43
  JOURNAL   Unpublished
```

AG151502/c

Pan troglodytes male lymphocytes DNA, clone_lib:RPCI-43 Chimpanzee
Male BAC Library clone:RP43-014K03.TJ

REFERENCE   2 (bases 1 to 664)
  AUTHORS   Fujiyama,A., Hattori,M., Toyoda,A., Taylor,T.D., Yada,T.,
            Totoki,Y., Watanabe,H. and Sakaki,Y.
  TITLE   Direct Submission
  JOURNAL   Submitted (02-AUG-2001) Asao Fujiyama, The Institute of Physical
            and Chemical Research (RIKEN), Genomic Sciences Center (GSC);
            1-7-22 Suehiro-chou,Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan
            (E-mail:chimpbes@gsc.riken.go.jp, URL:http://hgp.gsc.riken.go.jp/,
            Tel:81-45-503-9111, Fax:81-45-503-9170)
  COMMENT   Clones are derived from the chimpanzee BAC library RPCI-43 This BAC
            end was generated during the R&D process and may have higher chance
            of clone tracking errors.
            PRIMERS
              Sequencing: TJ
            LIBRARY
              Vector   : pBACe3.6
              R.Site 1 : EcoRI
              R.Site 2 : EcoRI.

FEATURES   Location/Qualifiers
     source   1..664
                     /organism="Pan troglodytes"
                     /db_xref="taxon:9598"
                     /clone="RP43-014K03.TJ"
                     /sex="male"
                     /cell_type="lymphocytes"
                     /clone_lib="RPCI-43 Chimpanzee Male BAC Library"
BASE COUNT   190 a   137 c   147 g   190 t
ORIGIN

Query Match   25.2%; Score 251.8; DB 17; Length 664;
Best Local Similarity   85.6%;
Matches 280; Conservative 0; Pred. No. 1.9e-26; Mismatches 47; Indels 0; Gaps 0;

```
Qy    467 TTGGCTGGGCACGGGTGGCTGCTCATGCCTGCAATCCCAGCACCTTGGGAGGCTGAGGCAGCA 526
Db    427 TCGGCTGGGCACGGTGCTCACTCCCTGTAATCCCAGCACTTTGGGAGGGCTGAGGCAGGTG 368
Qy    527 GATGACTTGAGCCAGGAGTTCGAGACCAGCCAGCCCAATATGGCGGAAACCCCATCTCATTTCTA 586
Db    367 GATCACTTGAGGTCAGGAGTTCAAGACCAGCCTGGCCAACATGGTGAAACCCCATCTCTA 308
Qy    587 CTAAAAATACAAAAAACTAACCAGGCGGTGGTGGCGGTGGCCTGTAGTCCCAGCTACTCGGG 646
Db    307 CTAAAAATACAAAAATTTAGCTGGGCGGTGGCCTGTAATCCCAGCTACCCGAGG 248
Qy    647 AGGCTGAGGCAGGAGAATCGGTTGAACCGGGAGGCAGAGGTTGCACTGAGCGTGAGATCA 706
Db    247 AGGCTGAGGCAGGAGAATCGCTTGAACCCGGGAGGCGGAGGTTGCAGTGAGCCGAGATCA 188
Qy    707 CACCACTGCACTCCAGCCTGGGCGACAGAGCGAGACTCCAGCTTAAAAAAAAAAAAAA 766
Db    187 CACCACTGCACTCCAGCCTGGGCGACAGAGCGAGACTCCAGCTTAAAAAAAAAAAAAAT 128
Qy    767 AAAAAAAAAAGGAGACCAATCACACTGCTGT 793
Db    127 AAAAAATAAACTGAATATTCCAGTGGT 101
```

  TITLE   and Venter,J.C.
            Use of a random BAC End Sequence Database for Sequence-Ready Map
            Building
  JOURNAL   Unpublished (1997)
  COMMENT   Other_GSSs: 345H13.TPB
            Contact: Mark Adams
            The Institute for Genomic Research
            9712 Medical Center Dr., Rockville, MD 20850, USA
            Tel: 301 838 0200
            Fax: 301 838 0208
            Email: mdadams@tigr.org
            Clones are available from Research Genetics (info@resgen.com). BAC
            end and search page:
            http://www.tigr.org/tdb/humgen/bac_end_search/bac_end_search.html
            Class: BAC ends.
            Seq primer: T7

FEATURES   Location/Qualifiers
     source   1..642
                     /organism="Homo sapiens"
                     /db_xref="GDB:5363319"
                     /db_xref="taxon:9606"
                     /clone="345H13"
                     /sex="Male"
                     /clone_lib="CIT-HSP"
                     /cell_type="sperm"
                     /note="Vector: pBeloBAC11; Site_1: HindIII; Site_2:
                     HindIII"
BASE COUNT   171 a   121 c   180 g   170 t
ORIGIN

Query Match   25.0%; Score 250.2; DB 17; Length 642;
Best Local Similarity   85.3%;
Matches 279; Conservative 0; Pred. No. 3.3e-26; Mismatches 48; Indels 0; Gaps 0;

```
Qy    454 CAAGAGACAGTCATTGGCGTGGGCGCACGGGTGGCTCATGCCTGCAATCCCAGCACCTTGGGAG 513
Db     38 CAAAAAATAGTTGTTGGCCGGGCACAGTGGCTCATGCCTGCAATCCCAGCACTTTGGGAG 97
Qy    514 GCTGAGGCAGGCAGATGACTTGAGGCCAGGAGTTCGAGACCAGCCTGGCCAATATGGCGA 573
Db     98 GCTAAGGCAGATGGATCACCTGAGGTCAGGAGTTCAAGACCAGCCTGGCCAACATGGTGA 157
Qy    574 AACCTCATTTCTACTAAAAATATACAAAAAACTAACCAGGCGTGGTGGCGGGCGCCTGTAATC 633
Db    158 AACCCCATCTCTACTAAAAATACAAAAAATTAGCCGGGCGTGGTGGCGCGCGCCTGTAATC 217
Qy    634 CCAGCTACTCGGGAGGCTGAGGCAGGAGAATCGCTTGAACCCGGGAGGCAGAGGTTGCAG 693
Db    218 CCAGCTACTCGGGAGGCTGAGGCAGGAGAATCGCTTGAACCCGGGAGGCGGAGGTTGCAG 277
Qy    694 TGAGCCGAGATCACACCACTGCACTCCAGCCTGGGCGACAGAGCGAGACTCCAGCTTAAA 753
Db    278 TGAGCCGAGATTGTGCCACTGCACTCCAGCCTGGGCAACAGAGCGAGACTCCATCTCAAAA 337
Qy    754 AAAAAAAAAAAAAAAAAAAAAAAAAAAGTAAAGAAA 780
Db    338 AAAAAAAAAAAAAAAAAAAAAAAAAAAGTAAAGAAA 364
```

RESULT 4
B59854
LOCUS   B59854   642 bp   DNA   linear   GSS 21-JUN-1998
DEFINITION CIT-HSP-345H13.TVB CIT-HSP Homo sapiens genomic clone 345H13, DNA
            sequence.
ACCESSION   B59854
VERSION   B59854.1 GI:2614572
KEYWORDS   GSS.
SOURCE   human.
  ORGANISM Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1 (bases 1 to 642)
  AUTHORS   Adams,M.D., Rounsley,S.D., Field,C.E., Bass,S., Linher,K., Golden
            ,K., Berry,K., Granger,D., Suh,E., Wible,C., Shizuya,H., Simon,M.

RESULT 5
BM802793
LOCUS   BM802793   1100 bp   mRNA   linear   EST 05-MAR-2002
DEFINITION AGENCOURT_6459887 NIH_MGC_92 Homo sapiens cDNA clone IMAGE:5575110
            5', mRNA sequence.
ACCESSION   BM802793
VERSION   BM802793.1 GI:19119616
KEYWORDS   EST.
SOURCE   human.
  ORGANISM Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1 (bases 1 to 1100)

```
AUTHORS     NIH-MGC http://mgc.nci.nih.gov/.
TITLE       National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL     Unpublished (1999)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: ATCC
            cDNA Library Preparation: Life Technologies, Inc.
            cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
            DNA Sequencing by: Agencourt Bioscience Corporation
            Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLAM12324 row: n  column: 07
            High quality sequence stop: 687.
FEATURES             Location/Qualifiers
     source          1..1100
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:5575110"
                     /clone_lib="NIH_MGC_92"
                     /tissue_type="embryonal carcinoma, cell line"
                     /lab_host="DH10B (phage-resistant)"
                     /note="Organ: testis; Vector: pCMV-SPORT6; Site_1: NotI;
                     Site_2: SalI; Cloned unidirectionally; oligo-dT primed.
                     Average insert size 2.5 kb. Library enriched for
                     full-length clones and constructed by Life Technologies.
                     Note: this is a NIH_MGC Library."
BASE COUNT      236 a    372 c    303 g    188 t      1 others
ORIGIN

Query Match             24.8%;  Score  248;  DB 14;  Length 1100;
  Best Local Similarity   76.7%;  Pred. No. 4.8e-26;
  Matches   330;  Conservative    0;  Mismatches   95;  Indels    5;  Gaps    2;

Qy         361 CCCTACCCAACTACCCCACCCCACCCCCAAGCCATGGCAGGCCCCAGGAACTTGATCCTGG 420
               |||| |||| || |||   ||| ||| | | ||| ||| | |   ||| |  | | ||
Db          84 CCCTAGCCAACGTGGACACCCCTCCGACGCGCGTGCCCCCTCAGAATATTCTCATTCCCT 143

Qy         421 GCTTTGCCGTATGCCAAGTCCTTACACCCCTCTCAAGAGA---CAGTCATTGGCTGGGCA 477
               | ||||||  |  |   |   | |  |||||||| | |||     ||||||||||| ||
Db         144 CCTTGCCCAAATATCCCCTCCCACTCTCATTAAGAGATACTCCCCACCTTGGCCGGGAG 203

Qy         478 CGGTGGCTCATGCCTGCCAATCCAGCACCTTGGGAGGCCGAGGCAGGAGGACTTGAG 537
               |||||||||| |||| ||ATCC ||||| |||||||||||||||||||||||   |||
Db         204 CGGTGGCTCACGCCTGTAATCCCAGCACTTTGGGAGGCCGAGGCAGGAGGATCACCTGAG 263

Qy         538 GCCAGGAGTTCGAGACCAGCCTGGCCAATATGGCGAAACCTCATTTCTACTAAAAATACA 597
               | |||||||||||||||||| |||||||||||||| |  |  |||  | |||||||||| |
Db         264 GTCAGGAGTTCGAGACCAGCAGCCTAGCCAACATGGTGAAACCCCATCTCTACTAAAAATACA 323

Qy         598 A--AAACTAACCAGGCGTGGTGGTGGTGTGCCTGGGCTGTTGCCTGTGGAGGCTGAGG 655
               |   |||| ||| ||||||||||  | | | | | ||  | | |||| ||||||||||||
Db         324 AAGAAAATTAGCCGGGCGTGGTGGCGGGCGCCTGTAATCCCAGCTACTCGGGAGGCTGAGG 383

Qy         656 CAGGAGAATCGCTTGAACCGGGAGGCGGAGGTTGCAGTGAGCGAGATCACACCACTGC 715
               ||||||||||||||||||| ||||| | |||||||||| |  ||||| |  || | ||
Db         384 CAGGAGAATCGCTTGAACCTGGAGGTTGCAGTGAGCCGAGATCGCGCCATTGC 443

Qy         716 ACTCCAGCCTGGGCGACAGAGCGAGACTCCAGCTTAAAAAAAAAAAAAAAAAAAA 775
               |||||||||| |||||||||||||||||||  | |||||||||||| ||||||||||||
Db         444 ACTCCAGCCTCGGCGACAGAGCGAGACTCCGTCTCAAAAAAAAAAAAGAAAAAAAAAAA 503

Qy         776 GGGAGACCATC 785
               | || | | ||
Db         504 AGATACTGTC 513
```

```
RESULT 6
ACCESSION   AQ415030
LOCUS       AQ415030      711 bp    DNA     linear   GSS 23-MAR-1999
DEFINITION  RPCI-11-203E3.TV RPCI-11 Homo sapiens genomic clone RPCI-11-203E3,
            DNA sequence.
ACCESSION   AQ415030
VERSION     AQ415030.1  GI:4473999
KEYWORDS    GSS.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 711)
  AUTHORS   Zhao,S., Adams,M.D., Nierman,W., Malek,J., de Jong,P. and Venter,
            J.C.
  TITLE     Use of BAC End Sequences from Library RPCI-11 for Sequence-Ready
            Map Building
  JOURNAL   Unpublished (1997)
COMMENT     Other GSSs: RPCI-11-203E3.TJ
            Contact: Shaying Zhao, William Nierman, Mark Adams
            Department of Eukaryotic Genomics
            The Institute for Genomic Research
            9712 Medical Center Dr., Rockville, MD 20850
            Tel: 301 838 0200
            Fax: 301 838 0208
            Email: hbe@tigr.org
            Clones are derived from the human BAC library RPCI-11. For BAC
            library availability, please contact Pieter de Jong
            (pieter@dejong.med.buffalo.edu). Clones may be purchased from
            BACPAC Resources (http://bacpac.med.buffalo.edu/ordering) or from
            Research Genet cs (info@resgen.com). BAC end search page:
            http://www.tigr.org/tdb/humgen/bac_end_search/bac_end_search.html.
            Seq_primer: T7
            Class: BAC ends.
FEATURES             Location/Qualifiers
     source          1..711
                     /organism="Homo sapiens"
                     /db_xref="GDB:7577666"
                     /db_xref="taxon:9606"
                     /clone="RPCI-11-203E3"
                     /clone_lib="RPCI-11"
                     /sex="Male"
                     /cell_type="Lymphocytes"
                     /note="Vector: pBACe3.6; Site_1: EcoRI; Site_2: EcoRI;
                     RPCI11 Human Male BAC Library"
BASE COUNT      270 a    150 c    141 g    150 t
ORIGIN

Query Match             24.7%;  Score  247;  DB 17;  Length 711;
  Best Local Similarity   85.0%;  Pred. No. 7.1e-26;
  Matches   277;  Conservative    0;  Mismatches   49;  Indels    0;  Gaps    0;

Qy         455 AAGAGACAGTCATTGGCTGGGCACAGGTGGCTCATGCCTGCAATCCCAGCGACCTTGGGAGG 514
               ||  | ||  | || | |||||||| |||||||||||||| || | |||| | | ||||||
Db          76 AAATGAAAAACTATAGGCCGGGCACAGTGGCTCATGCCTGTAATTCCAGCACTTTGGGAGG 135

Qy         515 CTGAGGCGAGGCAGATGACTTGAGGCCAGGAGTTCGAGACCAGCCTGGCCAATATGGCGA 574
               ||||||| | ||| || || ||||| ||||||||| ||||||||||| |||| |||| |
Db         136 CTGAGGCTGGCGGATCACCTGAGGTCAGGAGTTCAAGACCAGCCTGGCCAACATGGAGAA 195

Qy         575 ACCTCATTTCTACTAAAAATACAAAAAAATTAGCCAGGCGTGGTGGCGTTGCCTGTAATCC 634
               ||| | | |||||||||||||||||||||||| | ||||| | |||| || |||||||||
Db         196 ACCCCTCTCTACTAAAAATACAAAAAAATTAGCTGGGCATCGTGGGTGCACCTGTAATCC 255

Qy         635 CAGCTACTCGGGAGGCTGAGGCAGGAGAATCGCTTGAACCGGGGAGGCGGAGGTTGCAGT 694
               ||  |||||||||||||||| |||||||||  |||||||| ||| ||||||||||||||||
Db         256 CACCTACTCGGGAGGCTGAGACAGGAGAATTGCTTGAACCCGGGAGGCGGAGGTTGCAGT 315

Qy         695 GAGCTGAGATCACACCACTGCACTCCACCTCCGCCTGGGCGACAGAGCGAGACTCCAGCTTAAA 754
               |||| ||||||| |||||||||||||| || | |||||||||||||||| |||||||| | |||
Db         316 GAGCCGAGATCATGCCACTGCACTCCAGCCTGGCGACAGAGCAAGACTCCAATCTCAAA 375

Qy         755 AAAAAAAAAAAAAAAAAAAAAGGAGA 780
               |||||||||||||||||||||  | |
Db         376 AAAAAAAAAAAAAAAAAAAAAA 401
```

```
RESULT 7
AA515728/c
```

```
LOCUS       AA515728      439 bp    mRNA      linear    EST 19-AUG-1997
DEFINITION  ng70f04.s1 NCI_CGAP_Lip2 Homo sapiens cDNA clone IMAGE:940159
            similar to contains Alu repetitive element;, mRNA sequence.
ACCESSION   AA515728
VERSION     AA515728.1  GI:2255328
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 439)
  AUTHORS   NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
  TITLE     National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
            Tumor Gene Index
  JOURNAL   Unpublished (1997)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: L. Jeffrey Medeiros, M.D., Michael R.
            Emmert-Buck, M.D., Ph.D.
            cDNA Library Preparation: David B. Krizman, Ph.D.
            cDNA Library Arrayed by: Greg Lennon, Ph.D.
            DNA Sequencing by: Washington University Genome Sequencing Center
            Clone distribution: NCI-CGAP clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            www-bio.llnl.gov/bbrp/image/image.html
            Insert Length: 634   Std Error: 0.00
            Seq primer: pAMP10; mRNA made from liposarcoma, cDNA
            made by oligo-dT priming. Non- directionally cloned.
            Size-selected on agarose gel, average insert size 600 bp.
            Reference: Krizman et al. (1996) Cancer Research
            56:5380-5383."
            High quality sequence stop: 358.
FEATURES             Location/Qualifiers
     source          1..439
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:940159"
                     /clone_lib="NCI_CGAP_Lip2"
                     /tissue_type="liposarcoma"
                     /lab_host="DH10B"
                     /note="Vector: pAMP10;
BASE COUNT      121 a     89 c      95 g     134 t
ORIGIN
```

```
Query Match  24.7%;  Score 247.4;  DB 9;  Length 439;
Best Local Similarity  85.7%;
Matches 275;  Conservative 0;  Mismatches 46;  Indels 0;  Gaps 0;  Pred. No. 1e-25;

Qy    455 AAGAGACAGTCAATGGCTGGGCACGGGTGGCTCATGCCTGCAATCCCAGCACCTTGGGAGG 514
              |||||| |||| |||||||||||||||||||||||||| ||||||||||||| |||||||
Db    375 AAGAGAAACTAATAGGCTGGGCGTGGGTGGCTCATGCCTATAATCCCAGCACTTTGGGAGG 316

Qy    515 CTGAGGCAGGCAGATGACTTGAGGCCAGGAGTTCGAGACCAGCCTGGGCAACATGGCCGAA 574
              |||||||| |||| ||| ||||||| |||||||||| |||||||||||| ||||| ||||
Db    315 CTGAGGTCGGGCGGATCACTTGAGGTCAGGAGTTTGAGACCAGCCTGGGCAACATGGCGAA 256

Qy    575 ACCTCATTTCTACTAAAAATACAAAAAACTAACCAGGCGTGGTGGCTTGTTGCCTGTAATCC 634
              ||| ||| ||||||||||||||||||| |||||||||||||||||  |||||||||||
Db    255 ACCCCGGCTCTACTAAAAATACAAAAATTAGCCAGGCGTGGTGGTGTGCAGCTGTAATCC 196

Qy    635 CAGCTACTCCGGGAGGCTGAGGCAGGAGAATCGCTTGAACCCGGGAGGTTGCAGT 694
              |||||||| ||||||||||||||||||||||| ||||||||| |||||||||||
Db    195 CAGCTACTCGGGAGGCTGAGGCAGGAGAATCGCTTGAACCTGGGAGGTTGCAGT 136

Qy    695 GAGCTGAGATCACACACCTGCACTCCAGCCTGGGCGACAGAGCGAGACTCCATTTCAAAA 754
              ||||||||||| |||| |||||||||||||||||| |||||||||||||||| | ||||
Db    135 GAGCTGAGATCGTGCCACTGCATTCCAGCCTGGGTGACAGAGTGAGACTCCATTTCAAAA 76

Qy    755 AAAAAAAAAAAAAAAAAGAAAAGAAA 775
              ||||||||||||||||| |||||||
Db     75 AAAAAAAAAAAAAAAGAAAAGAAA 55
```

RESULT 8

```
LOCUS       B15692       460 bp    DNA       linear    GSS 04-JUN-1998
DEFINITION  345N8.TP CIT978SKA1 Homo sapiens genomic clone A-345N08, DNA
            sequence.
ACCESSION   B15692
VERSION     B15692.1  GI:2123441
KEYWORDS    GSS.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 460)
  AUTHORS   Adams,M.D., Kelley,J.M., Rounsley,S.R. and Venter,J.C.
  TITLE     Use of a BAC End Sequence Database for Sequence-Ready Map Building
  JOURNAL   Unpublished (1997)
COMMENT     Other_GSSs: 345N08.TV 345N08.TVB
            Contact: Mark Adams
            Department of Eukaryotic Genomics
            The Institute for Genomic Research
            9712 Medical Center Dr., Rockville, MD 20850, USA
            Tel: 301 838 0200
            Fax: 301 838 0208
            Email: mdadams@tigr.org
            Clones are available from Research Genetics (info@resgen.com). BAC
            end search page:
            http://www.tigr.org/tdb/humgen/bac_end_search/bac_end_search.html
            Seq primer: SP6
            Class: BAC ends.
FEATURES             Location/Qualifiers
     source          1..460
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="A-345N08"
                     /clone_lib="CIT978SKA1"
                     /sex="Female"
                     /cell_type="Fibroblast"
                     /note="Vector: pBAC108L; Site_1: HindIII; Site_2: HindIII;
                     CalTech Human BAC Library A1"
BASE COUNT      145 a     96 c     114 g     104 t
ORIGIN
```

```
Query Match  24.7%;  Score 247.2;  DB 17;  Length 460;
Best Local Similarity  87.7%;
Matches 270;  Conservative 0;  Mismatches 38;  Indels 0;  Gaps 0;  Pred. No. 1e-25;

Qy    468 TGGCTGGGCACGGGTGGCTCATGCCTGCAATCCCAGCACCTTGGGAGGCCGAGGCAGGTGG 527
              |||||||||| |||||||||||||||| |||||||||||| |||||||| |||||| |||
Db     74 TGGCTGGGCGCGGGTGGCTCATGCCTGTAATCCCAGCACTTTGGGAGGCCAAGGCAGGCGG 133

Qy    528 ATGACTTGAGGCCAGGAGTTCGAGACCAGCCTGGGCAACATGGCGAAACCTCATTTCTAC 587
              || |||||||| ||||||||| ||||||||||||||||||| |||| |||| ||||||||
Db    134 GTCACTTGAGGTCAGGAGTTCGAGACCAGCCTGGCCAACATGGCGAAACCCCGTCTCTAC 193

Qy    588 TAAAAATACAAAAAACTAACCAGGCGTGGTGGCGTGTGCCTGTAATCCCAGCTACTCGGGA 647
              ||||||||||||||  |||||||||||||||| ||| ||||||||||||||||||||||
Db    194 TAAAAATACAAAAATTAGCCAGGCGTGGTGGCGCGCGCCTGTAGTCCCAGCTACTCGGGA 253

Qy    648 GGCTGAGGCAGGAGAATCGCTTGAACCCGGGAGGTTGCAGTGAGCTGAGATCGCGAGATCAC 707
              ||||||||||||||||||||||| |||||||||||||||||||||||||||| ||||||||
Db    254 GGCTGAGGCAGGAGAATAGCTTGAACCCAGGAGGTTGCAGTGAGCCGAGATCGCGCCAGATCAC 313

Qy    708 ACCACTGCACTCCAGCCTGGGCGACAGAGCGAGACTCCAGCTTAAAAAAAAAAAAAAAAAA 767
              |||||||||||||||||||||| ||||||||||||||||| ||||||||||||||||||
Db    314 ACCACTGCACTCCAGCCTGGGCAACGAGAGCAAGAATCCGTCTCAAAAAAATAAAAAATAAAAAT 373

Qy    768 AAAAAAAA 775
              ||| |||
Db    374 AAAATAAA 381
```

RESULT 9

```
LOCUS       AA832145       430 bp    mRNA    linear   EST 05-MAR-1998
DEFINITION  oe22g09.s1 NCI_CGAP_Br5 Homo sapiens cDNA clone IMAGE:1386688
            similar to contains Alu repetitive element;contains element PTR7
            repetitive element ;, mRNA sequence.
ACCESSION   AA832145
VERSION     AA832145.1  GI:2905244
KEYWORDS    EST.
SOURCE      human.
ORGANISM    Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 430)
  AUTHORS   NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
  TITLE     National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
            Tumor Gene Index
  JOURNAL   Unpublished (1997)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: Ilan Kirsch, M.D., Kristina A. Cole, M.D.,
            Ph.D. student, Rodrigo F. Chuaqui, M.D., Michael R. Emmert-Buck,
            M.D., Ph.D.
            cDNA Library Preparation: David B. Krizman, Ph.D.
            cDNA Library Arrayed by: Greg Lennon, Ph.D.
            DNA Sequencing by: Washington University Genome Sequencing Center
            Clone distribution: NCI-CGAP clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            www-bio.llnl.gov/bbrp/image/image.html
            Insert Length: 554  Std Error: 0.00
            Seq primer: -40m13 fwd ET from Amersham.
FEATURES             Location/Qualifiers
     source          1..430
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:1386688"
                     /clone_lib="NCI_CGAP_Br5"
                     /sex="female"
                     /tissue_type="infiltrating ductal carcinoma"
                     /lab_host="DH10B"
                     /note="Organ: breast; Vector: pAMP10; mRNA made from
                     infiltrating ductal carcinoma, cDNA made by oligo-dT
                     priming. Non-directionally cloned. Size-selected on
                     agarose gel, average insert size 600 bp."
BASE COUNT    113 a    103 c    125 g     89 t
ORIGIN

Query Match             24.7%;  Score 246.8;  DB 9;  Length 430;
  Best Local Similarity   83.1%;
  Matches  281;  Conservative    0;  Mismatches   57;  Indels    0;  Gaps    0;

Qy        421 GCTTTGCGTATGCCAAGTCCTTACACCCCTCTCAAGAGACAGTCATTGGCTCGGCACGG 480
              |||||| ||||| || ||||| |  ||| || | || ||||| | ||||| || | |||
Db         93 GCTTTCCGTGTACCATGATTACTCTTCCTGGTCTTAATAGTCAGGGCCGTAGTCAGGGCGTGG 152

Qy        481 TGGCTCATGCCTGCAATCCAGCACCTTGGGAGGCCAGGCAGGCAGAATGACTTGAGGCC 540
              |||||| ||||| ||||| ||||| |||||||||||  |||||| || |||||||||| |
Db        153 TGGCTCACGCCTGTAATCCCAGCACTTTGGGAGGCCGAGGCAGGTAGATCACTTGAGGTC 212

Qy        541 AGGAGTTCGAGACCAGCCTGGCCAATATGGCGAAAACCTCATTTCTACTAAAAATACAAAA 600
              ||| ||||||||||||||||||||| ||||  |||| || || | |||||||||||||||
Db        213 AGGGGTTCGAGACCAGCCTGGCCAACATGGTGAAACCCCATCTGCTACTAAAAATACAAAA 272

Qy        601 ACTAACACAGGGCTGGGCGTGGTGGCTCGGCCTGTAATCCCAGCTACTCGGGAGGCAGGA 660
              | |  ||  ||| |||||||||||||| | |||||| ||||||||||| |||||||||||
Db        273 ATGAGCCGGGGCGTGGTGGCGCGTGCCTGTAGTCCCAGCTACTCAGGAGGCTGAGGCAGGA 332

Qy        661 GAATCGCTTGAACCAGGGAGGCGGAGGCAGCAGAGGTTGCAGTGAGCCACAACACTGCACTCC 720
              |||| || ||||| ||||||||||||| |  ||| |  ||||||||||||||||||
Db        333 GAATGGCGTGAACCCGGGAGGCGGAGGAGACTCCATCTCAAAAAAA 392

Qy        721 AGCCTGGGCGACAGAGCGAGACTCCAGCTTAAAAAAA 758
              |||||||||||||||||||||||||||||||||||||
Db        393 AGCCTGGGCGACAGAGCGAGACTCCAGCTTAAAAAAA 430
```

RESULT 10
BH367159/c

```
LOCUS       BH367159       416 bp    DNA    linear   GSS 03-DEC-2001
DEFINITION  UP_503-13K_T7 RPCI11 Human Male BAC Library Homo sapiens genomic
            clone 503-13K, DNA sequence.
ACCESSION   BH367159
VERSION     BH367159.1  GI:17297893
KEYWORDS    GSS.
SOURCE      human.
ORGANISM    Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 416)
  AUTHORS   Cheung,V.G., Dalrymple,H.L., Narasimhan,S., Watts,J., Schuler,G.,
            Raap,A.K., Morley,M. and Bruzel,A.
  TITLE     A resource of mapped human bacterial artificial chromosome clones
  JOURNAL   Genome Res. 9 (10), 989-993 (1999)
  MEDLINE   99455100
COMMENT     Contact: Arcaro MA, Morley M, Burdick J, Cheung VG
            Department of Pediatrics
            University of Pennsylvania
            3516 Civic Center Blvd, ARC 516, Philadelphia, PA 19104, USA
            Tel: 215 590 2664
            Fax: 215 590 3709
            Email: mlennox@mail.med.upenn.edu
            Seq primer: T7
            Plate: 503 row: K column: 13
            Class: BAC ends.
FEATURES             Location/Qualifiers
     source          1..416
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="503-13K"
                     /clone_lib="RPCI11 Human Male BAC Library"
                     /sex="Male"
                     /cell_type="Lymphocytes"
                     /note="Vector: pBACe3.6; RPCI11 Human Male BAC Library"
BASE COUNT     92 a    109 c     83 g    132 t
ORIGIN

Query Match             24.6%;  Score 246.6;  DB 17;  Length 416;
  Best Local Similarity   87.4%;
  Matches  270;  Conservative    0;  Mismatches   39;  Indels    0;  Gaps    0;

Qy        469 GGCTGGGCACGGTGGCTCATGCCTGCAATCCCAGCACTTTGGGAGGCCGAGGCAGGCAGA 528
              ||||||| | ||||||||||||||| |||||||||||||| | |||||||||| |||| |
Db        382 GGCTGGGTATGGTGGCTCATGCCTGTAATCCCAGCACTTTGAGAGGCCGAGGCGGGCACA 323

Qy        529 TGACTTGAGGCCAGGAGTTCGAGACCAGCCAGCGCCAACANTATGGCGAAACTCATTTCTACT 588
              | ||||| ||||| ||||||||||||||||  |||||||| | ||  |||| | || ||||
Db        322 TCACTTGGGGCCCAGGAGTTCGAGACCAGCCTGGCCAACANTGTCAAAACCCCTGTCTCTACT 263

Qy        589 AAAAATACAAAAAACTAAACCAGGCGTGGTGGTGGCGTGCCTGTAGTCCCAGCTACTCGGGAG 648
              ||||||||||||||  | || ||| |||||||   | |||||||| |||||||||||||||||
Db        262 AAAAATACAAAAAATTAGTTAGGCATGGTGGCAGGTGCCTGTAATCCCAGCTACTCGGGAG 203

Qy        649 GCTGAGGCAGGAGAATCGGCTTGAACGGGAGCGGAGGTTGCAGTGAGCCGAGATCACA 708
              ||||||||||||||||| ||||||| | ||| |||||||||||||||| ||||||| |
Db        202 GCTGAGGCAGGAGAATCGGCTTGAACCCGGAGGCGGAGGTTGCAGTGAGCTGAGATCATG 143

Qy        709 CCACTGCACTCCAGCCTGGGCGACAGAGCGAGACTCCAGCTTAAAAAAAAAAAAAAAA 768
              |||||||||||| ||||||| |||||||| |||||||| ||  ||| |||||||||||
Db        142 CCACTGCACTCCACCCTGGGTGACAGAGCAAGACTCCATCTCAAAATTAAAAAAAAAAAAA 83

Qy        769 AAAAAAAAGG 777
              ||||   ||
Db         82 AAAAGTAAG 74
```

RESULT 11
BG743198

```
LOCUS       BG743198      782 bp   mRNA   linear   EST 15-MAY-2001
DEFINITION  602634360F1 NCI_CGAP_Skn3 Homo sapiens cDNA clone IMAGE:4779326 5',
            mRNA sequence.
ACCESSION   BG743198
VERSION     BG743198.1  GI:14053851
KEYWORDS    EST.
SOURCE      human.
ORGANISM    Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 782)
AUTHORS     NIH-MGC http://mgc.nci.nih.gov/;
TITLE       National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL     Unpublished (1999)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: James Cleaver, M.D.
            cDNA Library Preparation: Life Technologies, Inc.
            cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL) DNA
            Sequencing by: Incyte Genomics, Inc.
            Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLAM10636  row: h  column: 15
            High quality sequence stop: 520.
FEATURES             Location/Qualifiers
     source          1..782
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:4779326"
                     /clone_lib="NCI_CGAP_Skn3"
                     /lab_host="DH10B (T1 phage-resistant)"
                     /note="Organ: skin; Vector: pCMV-SPORT6; Site_1: NotI;
                     Site_2: SalI; Cloned unidirectionally. Primer: Oligo dT.
                     Average insert size 1.5kb. Library constructed by Life
                     Technologies. Note: this is a NCI_CGAP Library."
BASE COUNT      230 a    196 c    184 g    172 t
ORIGIN

Query Match             24.68%;  Score 246.6;  DB 12;  Length 782;
Best Local Similarity   82.7%;  Pred. No. 9.3e-26;
Matches  282;  Conservative  0;  Mismatches  59;  Indels  0;  Gaps  0;

Qy    440 CCTTACACCCTCTCAAGAGACAGAGTCATTGGCGTGGGCACGGTGGCTGGCCTGCCAATCC 499
          ||| ||| |||||  ||||  |||| |||||  ||||||||||||||  ||   || |||
Db    171 CATTTTACAGAGTGTAGAAACTGAGGGCCAGGTGGCTCACGCCTTGAGTAGTCC       230

Qy    500 CAGCACCTTGGGAGGCCAGGCAGGCAGATGACTTGAGGCCAGAGTTCGAGACCAGCCT    559
          |||||||||||| |||||||| |||||  |||||  |||||| |||||||||||||||
Db    231 CAGCACTTTGGGGAGGCCAGGCAGGCAGATTAACCTGAGAGTTCTTGACCAGCCT          290

Qy    560 GGGCCAATATGGCGAAACCTCATTTCTACTAAAAAAATACAAAAAAACTAACCAGCCGTGGTGGC 619
          |||||  |   |||||   |   ||||||  |||||||||||||||||  |||||| ||| |||
Db    291 GGCCAACATGGTAAAAACCCATCTCTACTAAAAATACAAAAAATTAGCCAGGCATGGTGGC        350

Qy    620 TTGTGCCTGTAATCCCAGCTACTCGGGAGGCAGGAGAATCGCTTGAACCGGGGA    679
          ||||||||||||||||||||||| ||||||||||| |||||||||||| ||||
Db    351 ACGTCCCTGTAATCCCAGCTACTCGGGAGGCTGAGGCAGGAGAATCGCTTGAACCTGGGA      410

Qy    680 GGCGAGAGGTTGCAGTGAGCCGAGATCACACCACTGCACTCCAGCCTGGGCGACAGAGCGA    739
          |||||||| || | ||||| | |||  |||| ||||| | || ||||| ||||||| |||
Db    411 GGCAGAGGCTGCAGTGAGCCGAGATCGCACGCCACTCCAGCCTGGGCGACAGAGCGA        470

Qy    740 GACTCCAGCTTAAAAAAAAAAAAAAAAAAAAAAAAAAAAGGAGA   780
          ||| ||||||||||||||||||||||||||||||||||  | |
Db    471 GACTGTGTGTCTAAAAAAAAAATAAAAAAATAAAAAAAAAAAAAA      511
```

```
RESULT 12
LOCUS       AL704074      530 bp   mRNA   linear   EST 22-MAR-2002
DEFINITION  DKFZp68601228_r1 686 (synonym: hlcc3) Homo sapiens cDNA clone
            DKFZp68601228 5', mRNA sequence.
```

```
ACCESSION   AL704074
VERSION     AL704074.1  GI:19687429
KEYWORDS    EST.
SOURCE      human.
ORGANISM    Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 530)
AUTHORS     Wiemann,S.
            Ottenwaelder,B., Obermaier,B., Mewes,W., Mewes,H.W., Well,B. and
            Wiemann,S.
TITLE       EST (Ottenwaelder,B., Obermaier,B., Mewes,H.W., Well,B. and Wiemann
            ,S.)
JOURNAL     Unpublished (2001)
COMMENT     Contact: Ottenwaelder B
            MIPS
            Am Klopferspitz 18a D-82152 Martinsried, Germany
            This is the 5' sequence of the clone insert
            Clone from S. Wiemann, Molecular Genome Analysis, German Cancer
            Research Center (DKFZ); Email s.wiemann@dkfz-heidelberg.de;
            sequenced by MediGenomix (Martinsried/Germany) within the cDNA
            sequencing consortium of the German Genome Project. No s1 sequence
            available.
            This clone (DKFZp68601228) is available at the RZPD in Berlin.
            Please contact the R2PD: Ressourcenzentrum, Heubnerweg 6, 14059
            Berlin- Charlottenburg, GERMANY; Email: clone@rzpd.de.
FEATURES             Location/Qualifiers
     source          1..530
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="DKFZp68601228"
                     /clone_lib="686 (synonym: hlcc3)"
                     /tissue_type="human skeletal muscle"
                     /dev_stage="adult"
                     /lab_host="DH10B"
                     /note="Vector: pTriplEx2; Site_1: SfiIA; Site_2: SfiIB;
                     cDNA-collection"
BASE COUNT      116 a    130 c    101 g    183 t
ORIGIN

Query Match             24.68%;  Score 246.4;  DB 9;  Length 530;
Best Local Similarity   81.2%;  Pred. No. 1.2e-25;
Matches  286;  Conservative  0;  Mismatches  66;  Indels  0;  Gaps  0;

Qy    466 AATTGGGCTGGGGCACGGTGGCTCATGGCCTGCAATCCCAGCACCTTGGGAGGCTGAGGC 525
          ||||| |||||||||||||| ||| | |||||||||| ||||||||||||||| | |||
Db    455 ATTGGGGCCAGGCACGGTGGCTCACGCCTCGTAATCCCAGCACTTTGGGAGGCCGAGGC      396

Qy    526 AGATGACTTGAGGCCAGGAGTTCGAGACCAGCCTGGCCAATATGGTGGGCAAACCTCATTTCT 585
          ||| |||||||| |||||||||||||||| |||| | |||| |||||| |||  || || |
Db    395 AGATCACCTGAGGTCAGGAGTTCAAGACCAGCCTGGCCAACATGGTGAAACCCCATCTCT        336

Qy    586 ACTAAAAATACAAAAATTAACCAGGCGCGTGGTGGCGTGTGCCTGTAATCCCAGCTACTCGG 645
          ||||||| ||||||  |||||||| || |||||| |||| |||||||||||||||||||||
Db    335 ACTAAAAAACACAAAAATTAGCTGGGCGTGGTGGCGTGCGCCTGTAATCCCAGCTACTCAG      276

Qy    646 GAGGCTGAGGCAGGAGAAATCCCTTGAACCCGGGAGGCGGAGGTTGCAGTGAGCCGAGATC 705
          |||||| ||||||||||| |||||| ||||| ||||| |||||||||||||||||||||
Db    275 GAGGCTGAGGCAGGAGAATCGCTTGAACCTGGGAGGCAGAGGTTGCAGTGAGCCGAGATT      216

Qy    706 ACACCACTGCACTCCAGCCTGGGCGACAGAGCGAGACTCCAGCTTAAAAAAAAAAAAAAAA 765
          ||||| ||||||||||| ||| ||||||||| | |||  |||| ||||||||||||||||
Db    215 GCGCCATTGCACTCCAGCCTGGGCGACAGAGCAAGATTCCATCTCAAAAAAAAAAAAAAA      156

Qy    766 AAAAAAAAAAAGGAGACCAATCACTGCTGTCCTCCCATTTCTTACAGATGAAAAAA 817
          |||||||||| |||| ||   |||| |  |  || ||| || || ||||  ||||
Db    155 AAAATCAGTGAGATCACTATAGAGCCTTCTGACCTCCTCGAATGGTTTTAAAA       104
```

```
RESULT 13
LOCUS       B65134        657 bp   DNA   linear   GSS 21-JUN-1998
DEFINITION  CIT-HSP-2017C7.TRB CIT-HSP Homo sapiens genomic clone 2017C7, DNA
            sequence.
```

```
ACCESSION   B65134
VERSION     B65134.1  GI:2639112
KEYWORDS    GSS.
SOURCE      human.
ORGANISM    Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 657)
AUTHORS     Adams,M.D., Rounsley,S.D., Field,C.E., Bass,S., Linher,K., Golden
            ,K., Berry,K., Granger,D., Suh,E., Wible,C., Shizuya,H., Simon,M.,
            and Venter,J.C.
TITLE       Use of a random BAC End Sequence Database for Sequence-Ready Map
            Building
JOURNAL     Unpublished (1997)
COMMENT     Contact: Mark Adams
            Department of Eukaryotic Genomics
            The Institute for Genomic Research
            9712 Medical Center Dr., Rockville, MD 20850, USA
            Tel: 301 838 0200
            Fax: 301 838 0208
            Email: mdadams@tigr.org
            Clones are available from Research Genetics (info@resgen.com). BAC
            end search page:
            http://www.tigr.org/tdb/humgen/bac_end_search/bac_end_search.html
            Seq primer: M13 Reverse
            Class: BAC ends.
FEATURES             Location/Qualifiers
     source          1..657
                     /organism="Homo sapiens"
                     /db_xref="GDB:7043769"
                     /db_xref="taxon:9606"
                     /clone="2017C7"
                     /clone_lib="CIT-HSP"
                     /sex="Male"
                     /cell_type="Sperm"
                     /note="Vector: pBeloBAC11; Site_1: HindIII; Site_2:
                     HindIII"
BASE COUNT      219 a    128 c    143 g    167 t
ORIGIN

Query Match             24.6%;  Score 246;  DB 17;  Length 657;
  Best Local Similarity   76.9%;  Pred. No. 1.2e-25;
  Matches  300;  Conservative    0;  Mismatches   90;  Indels    0;  Gaps    0;

Qy        462 AGTCAATGGCGTGGGCACGGTGGCCTCACATGCCGCCAGCAGCTTGGGAGGCTCGAGGC 521
              ||| |||| |||||||| ||||||||| | |||||  ||| ||||||| ||||| |||
Db        141 AATACTTGGCGTGGGCAATGGGTGGCCTTATGCCTGTAAATCCACCACTTTGGAAGCTCGAGGT 200

Qy        522 AGGCAGATGACTTGAGGCCAGGAGTTCGAGACCAGCAGCCTGGCCAATATGGCGAAACCTCAT 581
              ||||  |||| | |||| |||||||||  | |||||  || ||| |||||  |||||| ||
Db        201 AGGCGGATCACCTGAAGTCAGGAGTTCAGGACCAGCGGCCCAACATGGTGAAACCTCAT 260

Qy        582 TTCTACTAAAAATACAAAAAACTAACCAGGTGTGGTGGCTGGGCGTAATCCCAGCTAC 641
              |||||||||||||||||||||  |  |||||| |||| | | || |||||||||||||
Db        261 CTCTACTAAAAATACAAAAAATTAGCCGGGTGGTGGTGCGTGTGCCTGTAATCCCAGCTAC 320

Qy        642 TCGGGAGGCTGAGGCAGGAGAATCGCTTGAACCGGGAGGTTGCAGTGAGCTGA 701
              | |||||||||||||||||||| ||||| || | |||||||||||||| |||
Db        321 TTGGGAGGCTGAGGCAGGAGAACTGCTTAAACCCAGGAGTTGCAGTGAGGTGA 380

Qy        702 GATCACCACCACTGCACTCCAGCCTGGGCGACAGAGCGAGACTCCAGCTTAAAAAAAAAA 761
              |||| ||| |||||||| ||||||||||||||||| |||| | ||| | ||||||| | |
Db        381 GATCCGCCGCCACTGCATTCCAGCCTGGGCGACAGAGTGAGAAACCAATCTCCAAAAAAACAAA 440

Qy        762 AAAAAAAAAAAAAGGAGAACGAACCACATCACTGCGTCCTCGCCATTCTTACAGATGAAAAACAGG 821
              ||||     |||  |||| | | || ||| || | ||| | |||||| |||  ||||||  |
Db        441 CAAACAAACAAACTTAAGTGCCAAACTTAACTAAGTAAATGCAGGAAT 500

Qy        822 CTCAGAGGTTGAATGCGTTTTCCTGAAGTCA 851
              || | | |||||| |   ||||  |||||||
Db        501 ACAAAAAAACGGGAAAGATATTCTGACATAA 530
```

```
RESULT 14
AW970571
LOCUS       AW970571     521 bp    mRNA    linear   EST 01-JUN-2000
DEFINITION  EST382652 MAGE resequences, MAGK Homo sapiens cDNA, mRNA sequence.
ACCESSION   AW970571
VERSION     AW970571.1  GI:8160416
KEYWORDS    EST.
SOURCE      human.
ORGANISM    Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 521)
AUTHORS     Hegde,P., Qi,R., Abernathy,K., Dharap,S., Gaspard,R., Gay,C., Holt
            ,I.E., Saeed,A.I., Sharov,V., Lee,N.H., Yeatman,T.J. and
            Quackenbush,J.
TITLE       Assessment of gene expression patterns in a model of colon tumor
            metastasis using a 19,200 element cDNA microarray
JOURNAL     Unpublished (2000)
COMMENT     Contact: John Quackenbush
            The Institute for Genomic Research
            9712 Medical Center Dr., Rockville, MD 20850, USA
            Tel: 301 838 3528
            Fax: 301 838 0208
            Email: johnq@tigr.org
            Plate: 279
            Seq primer: Forward.
FEATURES             Location/Qualifiers
     source          1..521
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="MAGE resequences"
                     /clone_lib="MAGE resequences, MAGK"
                     /note="Vector: pbluescriptSKm"
BASE COUNT      162 a    128 c    141 g     89 t      1 others
ORIGIN

Query Match             24.6%;  Score 245;  DB 10;  Length 521;
  Best Local Similarity   83.1%;  Pred. No. 1.5e-25;
  Matches  280;  Conservative    0;  Mismatches   57;  Indels    0;  Gaps    0;

Qy        444 ACACCCCTCTCAAGACAGACAGTCATTGGCGTGGCACGGTGGCCTCACATGCCGCCAATCCCAGC 503
              |  |    |  |||  |  | || |||||| |||| ||||||| | ||| |||||| || |||
Db        155 AGATCTCTGCAAAAGAAAATACCTTCCGGGCCGGGCGGCACAGTGGCCCTCGTAATCCCAGC 214

Qy        504 ACCTTGGGAGGCTGAGGCAGGCAGAGATGACTTGAGGCCAGGAGTTCGAGACCAGCCGCC 563
              | |||||||| |  | |||||  | || || || || || ||||| || | || | |||
Db        215 ATTTTTGGGAGAGGCCCCAGGCCGGGTGGATCACGAGGTCAGGAGATCACCTGAGGTCAGGAGT 274

Qy        564 AATATGGCGAAACCTCATTTCTACTAAAAATACAAAAATACAAAAATTACCAGGGCGGTGGCTTGT 623
              || ||||||||||| || | || |||||||||||||| |||||| || || |||| |||| |
Db        275 AACATGGCGAAACCCCGTGTCTCTACTAAAAATACAAAAATTAGCCGGGCGGTGGTGGCGGC 334

Qy        624 GCCTGTAATCCCAGCTACTCGGGAGGCTGAGGCAGGAGAATCGCTTGAACCGGGAGGCA 683
              |||||||||||  |||||| |||||||||||||||||||||| |||| |||| |||||
Db        335 GCCTGTAATCCACCAACTCGGGAGGCTGAGGCAGGAGAATCACTTGAGCCGGGAGGCA 394

Qy        684 GAGGTTGCAGTGAGCTGAGATCACACACCACTGCACTCCAGCCTGGGCGACACAGAGCGAGACT 743
              ||| |||||||||| | || || | | |||||||||||||||||||||||| |||||||||||
Db        395 GAGCTTGCAGTGAGCCGAGATCGCGCCACCACTGCACTCCAGCCTGGGCGACAGAGCGAGACT 454

Qy        744 CCAGCTTAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAGGAGA 780
              || |  | ||||  |||||||||||||||||| | | ||||
Db        455 CCGGTCTCAAAACAAAAAAAAAAAAAAAAAAAGAAAAAAAA 491

RESULT 15
BM993322/c
LOCUS       BM993322     552 bp    mRNA    linear   EST 17-JUN-2002
DEFINITION  UI-H-DT0-avi-b-08-0-UI.s1 NCI_CGAP_DT0 Homo sapiens cDNA clone
            IMAGE:5879719 3', mRNA sequence.
ACCESSION   BM993322
VERSION     BM993322.1  GI:19712767
KEYWORDS    EST.
```

```
SOURCE      human.
ORGANISM    Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 552)
AUTHORS     NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
TITLE       National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
            Tumor Gene Index
JOURNAL     Unpublished (1997)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: Dr. Jose Mercuende
            cDNA Library Preparation: Dr. M. Bento Soares, University of Iowa
            cDNA Library Arrayed by: Dr. M. Bento Soares, University of Iowa
            DNA Sequencing by: Dr. M. Bento Soares, University of Iowa
            Clone Distribution: Clone distribution information can be found
            through the I.M.A.G.E. Consortium/LLNL at: http://image.llnl.gov
            The following repetitive elements were found in this cDNA
            sequence: 11-300, >AlU (matched compliment)
            Seq primer: M13 FORWARD
            POLYA=Yes.
FEATURES             Location/Qualifiers
     source          1..552
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone_lib="NCI_CGAP_DT0"
                     /clone="IMAGE:5879719"
                     /tissue_type="Metastatic Chondrosarcoma"
                     /dev_stage="Adult"
                     /lab_host="DH10B (Life Technologies)"
                     /note="Organ: Lung; Vector: pT7T3-Pac (Pharmacia) with a
                     modified polylinker; Site_1: EcoR I; Site_2: Not I;
                     NCI_CGAP_DT0 is a cDNA library containing the following
                     tissue(s): Metastatic Chondrosarcoma in Lung. The library
                     was constructed according to Bonaldo, Lennon and Soares,
                     Genome Research, 6:791-806, 1996. First strand cDNA
                     synthesis was primed with an oligo-dT primer containing a
                     Not I site. Double stranded cDNA was ligated to an EcoR I
                     adaptor, digested with Not I, and cloned directionally
                     into pT7T3-Pac vector. The oligonucleotide used to prime
                     the synthesis of first-strand cDNA contains a library tag
                     sequence that is located between the Not I site and the
                     (dT)18 tail. The sequence tag for this library is
                     AACTGTTCGG.
                     TAG_LIB=UI-H-DT0
                     TAG_TISSUE=lung metastatic chondrosarcoma
                     TAG_SEQ=AACTGTTCGG"
BASE COUNT       123 a    111 c    111 g    177 t
ORIGIN
```

```
Query Match             24.5%;  Score 245.6;  DB 14;  Length 552;
Best Local Similarity   86.1%;
Matches   272;  Conservative   0;  Mismatches   44;  Indels   0;  Gaps   0;

Qy    453 TCAAGAGACAAGTCATTGGCTGGGCACGGTGGCTCAATCCCAGCACCTTGGGA 512
          | ||| ||   | ||||||||  | ||||||||||||||||||||||||||||
Db    316 TGAAAAGACAAAAGCGGCTGGGCTTGGTGGCTCACACCTGTAATCCCAGCACTTTGGGA 257

Qy    513 GGCTGAGGCAGGCGAGATGACTTGAGGCCAGGAGTTCGAGACCAGCCTGGCCAATATGCCG 572
          |||||||||||| || ||| | |||| ||||| |||||||||||||||| |||||| ||
Db    256 GGCTGAGGCAGGCGAGATCACCTGAGGTCAGGAGTTCGAGACCAGCCTGACCTATATGGTG 197

Qy    573 AAACCTCATTTCTACTAAAAATACAAAAAACTAACCAGGCGTGGTGGCGTTGTGCCTGTAAT 632
          |||| | |  |||||| |   | |||||||||||| ||||||| |  ||||||||||||||
Db    196 AAACCCGTCTCTACTAAGAATACAAAAATTACCTGACCATGGTGGCAGGTGCCTGTAAT 137

Qy    633 CCCAGCTACTCGGGAGGCTGAGGCAGGAGAATCGCTTGAACCCGGGAGGCAGAGGTTGCA 692
          |||||||||||||||||||||||| |||||| |||||||||||||||||||||||||||
Db    136 TCCAGCTACTCGAGAGGCTGAGGACAGGAGAATCGCTTGAACCCGGGAGGCAGAGGTTGCA 77

Qy    693 GTGAGCTGAGATCACACCACTGCACTCGACTCCAGCCTGGGCGACAGAGCGAGACTCCAGCTTAA 752
          |  ||| |||||||  |||||||||||||||||||||||||||||||||||||||| || ||
Db     76 GCAAGCCAAGATCGTGCCACTGCACTCCAGCCTGGGCGACAGAGCGAGACTCCATCTCAA 17
```

THIS PAGE BLANK (USPTO)

OM nucleic - nucleic search, using sw model

GenCore version 5.1.4_p5_4578
Copyright (c) 1993 - 2003 Compugen Ltd.

Run on: March 29, 2003, 16:26:54 ; Search time 46.4809 Seconds

Title: US-09-691-219-3_COPY_10000_11000
Perfect score: 1001
Sequence: 1 ttactttaccatcctgggg.........ggccatgggtgaggacatg 1001
Scoring table: IDENTITY_NUC
    Gapop 10.0 , Gapext 1.0

Searched: 441362 seqs, 153338381 residues

Total number of hits satisfying chosen parameters: 882724

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
    Maximum Match 100%
    Listing first 45 summaries

Database :   Issued_Patents_NA:*
   1: /cgn2_6/ptodata/1/ina/5A_COMB.seq:*
   2: /cgn2_6/ptodata/1/ina/5B_COMB.seq:*
   3: /cgn2_6/ptodata/1/ina/6A_COMB.seq:*
   4: /cgn2_6/ptodata/1/ina/6B_COMB.seq:*
   5: /cgn2_6/ptodata/1/ina/PCTUS_COMB.seq:*
   6: /cgn2_6/ptodata/1/ina/backfiles1.seq:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	243	24.3	162450	1	US-09-345-882-1	Sequence 1, Appli
2	240.6	24.0	7676	2	US-08-451-777A-7	Sequence 7, Appli
3	240.6	24.0	7676	2	US-08-451-778A-7	Sequence 7, Appli
4	240.6	24.0	7676	2	US-08-998-208-7	Sequence 7, Appli
5	240.6	24.0	7676	5	PCT-US95-06743-7	Sequence 7, Appli
6	238.8	23.9	162450	1	US-09-345-882-1	Sequence 1, Appli
7	237.2	23.7	4038	4	US-08-969-125-8	Sequence 8, Appli
8	237.2	23.7	62804	4	US-09-800-960-3	Sequence 3, Appli
9	237	23.7	29629	4	US-09-729-995-3	Sequence 3, Appli
10	236.8	23.7	70000	4	US-09-851-896-3	Sequence 3, Appli
11	236.4	23.6	1701	4	US-09-078-294-9	Sequence 9, Appli
12	236.2	23.6	43950	1	US-09-735-934A-3	Sequence 3, Appli
13	236.2	23.6	6769	1	US-08-480-784-20	Sequence 20, Appl
14	236.2	23.6	6769	1	US-08-483-553-20	Sequence 20, Appl
15	236.2	23.6	6769	1	US-08-487-002-20	Sequence 20, Appl
16	236.2	23.6	6769	1	US-08-483-554B-20	Sequence 20, Appl
17	236.2	23.6	6769	1	US-08-488-011B-20	Sequence 20, Appl
18	236.2	23.6	6769	4	US-08-850-727-20	Sequence 20, Appl
19	236.2	23.6	6769	5	PCT-US95-10202-20	Sequence 20, Appl
20	236.2	23.6	6769	5	PCT-US95-10203-20	Sequence 20, Appl
21	236.2	23.6	8453	5	PCT-US95-10220-20	Sequence 20, Appl
22	236	23.6	8453	2	US-09-167-681-45	Sequence 45, Appl
23	235.8	23.6	1988	2	US-08-257-963B-11	Sequence 11, Appl
24	235.8	23.6	1988	4	US-08-367-841A-11	Sequence 11, Appl
25	235.8	23.6	1988	5	PCT-US95-07201-11	Sequence 11, Appl
26	235.8	23.6	5262	4	US-08-520-373D-5	Sequence 5, Appli
27	235.8	23.6	22481	4	US-08-367-841A-43	Sequence 43, Appl
28	235.8	23.6	22481	5	PCT-US95-07201-43	Sequence 43, Appl
29	235.8	23.6	22484	4	US-09-875-223-2	Sequence 2, Appli
30	235.4	23.5	15602	4	US-09-844-634-17	Sequence 17, Appl
31	235.2	23.5	619	4	US-09-385-982-358	Sequence 358, App
32	235.2	23.5	631	4	US-09-385-982-354	Sequence 354, App
33	235.2	23.5	80595	4	US-09-078-294-3	Sequence 3, Appli
34	235	23.5	14581	4	US-08-520-373D-4	Sequence 4, Appli
35	235	23.5	80246	4	US-09-078-294-4	Sequence 4, Appli
36	234.6	23.4	12047	2	US-09-022-461-1	Sequence 1, Appli
37	234.6	23.4	12047	4	US-09-033-556-3	Sequence 3, Appli
38	234.6	23.4	50000	4	US-09-146-053-3	Sequence 3, Appli
39	234.2	23.4	36711	4	US-09-301-665-3	Sequence 3, Appli
40	234.2	23.4	43950	4	US-09-735-934A-3	Sequence 3, Appli
41	234	23.4	32042	4	US-09-245-281-44	Sequence 44, Appl
42	233.8	23.4	62804	4	US-09-800-960-3	Sequence 3, Appli
43	233.6	23.4	14636	4	US-09-173-914-6	Sequence 6, Appli
44	233.6	23.3	45546	4	US-09-146-053-6	Sequence 6, Appli
45	233.6	23.3	81001	4	US-09-750-580-1	Sequence 1, Appli

ALIGNMENTS

RESULT 1
US-09-345-882-1

Sequence 1, Application US/09345882
Patent No. 6393373
GENERAL INFORMATION:
APPLICANT: Bougueleret, Lydie
TITLE OF INVENTION: A NUCLEIC ACID ENCODING A RETINOBLASTOMA BINDING PROTEIN (RBP-
TITLE OF INVENTION: AND POLYMORPHIC MARKERS ASSOCIATED WITH SAID NUCLEIC ACID.
FILE REFERENCE: GENSET.031A
CURRENT APPLICATION NUMBER: US/09/345,882
CURRENT FILING DATE: 1999-06-30
PRIOR APPLICATION NUMBER: US 60/091,315
PRIOR FILING DATE: 1998-06-30
PRIOR APPLICATION NUMBER: US 60/111,909
PRIOR FILING DATE: 1998-12-10
NUMBER OF SEQ ID NOS: 140
SOFTWARE: Patent.pm

SEQ ID NO 1
LENGTH: 162450
TYPE: DNA
ORGANISM: Homo sapiens
FEATURE:
NAME/KEY: allele
LOCATION: 72794
OTHER INFORMATION: 5-124-273 : polymorphic base A or G
FEATURE:
NAME/KEY: allele
LOCATION: 88073
OTHER INFORMATION: 5-127-261 : polymorphic base A or C
FEATURE:
NAME/KEY: allele
LOCATION: 90842
OTHER INFORMATION: 99-1437-325 : polymorphic base A or G
FEATURE:
NAME/KEY: allele
LOCATION: 93714
OTHER INFORMATION: 5-128-60 : polymorphic base deletion of GT
FEATURE:
NAME/KEY: allele
LOCATION: 97122
OTHER INFORMATION: 99-1442-224 : polymorphic base G or T
FEATURE:
NAME/KEY: allele
LOCATION: 97152
OTHER INFORMATION: 5-129-144 : polymorphic base deletion of T
FEATURE:
NAME/KEY: allele
LOCATION: 99098
OTHER INFORMATION: 5-130-257 : polymorphic base A or G
FEATURE:

FEATURE:
NAME/KEY: allele
LOCATION: 9917
OTHER INFORMATION: 5-130-276 : polymorphic base A or G

FEATURE:
NAME/KEY: allele
LOCATION: 103806
OTHER INFORMATION: 5-131-395 : polymorphic base A or T

FEATURE:
NAME/KEY: allele
LOCATION: 106940
OTHER INFORMATION: 5-133-375 : polymorphic base insertion of A

FEATURE:
NAME/KEY: allele
LOCATION: 108106
OTHER INFORMATION: 5-135-155 : polymorphic base insertion of A

FEATURE:
NAME/KEY: allele
LOCATION: 108149
OTHER INFORMATION: 5-135-198 : polymorphic base insertion of GTTT

FEATURE:
NAME/KEY: allele
LOCATION: 108308
OTHER INFORMATION: 5-135-357 : polymorphic base A or G

FEATURE:
NAME/KEY: allele
LOCATION: 108471
OTHER INFORMATION: 5-136-174 : polymorphic base C or T

FEATURE:
NAME/KEY: allele
LOCATION: 134134
OTHER INFORMATION: 5-140-120 : polymorphic base C or T

FEATURE:
NAME/KEY: allele
LOCATION: 134362
OTHER INFORMATION: 5-140-348 : polymorphic base insertion of A

FEATURE:
NAME/KEY: allele
LOCATION: 13374
OTHER INFORMATION: 5-140-361 : polymorphic base insertion of CA

FEATURE:
NAME/KEY: allele
LOCATION: 146328
OTHER INFORMATION: 5-143-84 : polymorphic base insertion of A

FEATURE:
NAME/KEY: allele
LOCATION: 146345
OTHER INFORMATION: 5-143-101 : polymorphic base A or C

FEATURE:
NAME/KEY: allele
LOCATION: 150329
OTHER INFORMATION: 5-145-24 : polymorphic base A or G

FEATURE:
NAME/KEY: allele
LOCATION: 160031
OTHER INFORMATION: 5-148-352 : polymorphic base G or T

FEATURE:
NAME/KEY: allele
LOCATION: 72771..72817
OTHER INFORMATION: polymorphic fragment 5-124-273 SEQ ID30

FEATURE:
NAME/KEY: allele
LOCATION: 72771..72817
OTHER INFORMATION: polymorphic fragment 5-124-273 SEQ ID31

FEATURE:
NAME/KEY: allele
LOCATION: 88050..88096
OTHER INFORMATION: polymorphic fragment 5-127-261 SEQ ID51

FEATURE:
NAME/KEY: allele
LOCATION: 88050..88096
OTHER INFORMATION: polymorphic fragment 5-127-261 SEQ ID52

FEATURE:
NAME/KEY: allele
LOCATION: 90819..90865
OTHER INFORMATION: complement polymorphic fragment 99-1437-325 SEQ ID49

FEATURE:
NAME/KEY: allele
LOCATION: 90819..90865
OTHER INFORMATION: complement polymorphic fragment 99-1437-325 SEQ ID70

FEATURE:
NAME/KEY: allele
LOCATION: 93690..93736
OTHER INFORMATION: polymorphic fragment 5-128-60 SEQ ID53

FEATURE:
NAME/KEY: allele
LOCATION: 93690..93736
OTHER INFORMATION: polymorphic fragment 5-128-60 SEQ ID32

FEATURE:
NAME/KEY: allele
LOCATION: 97099..97145
OTHER INFORMATION: polymorphic fragment 99-1442-224 SEQ ID50

FEATURE:
NAME/KEY: allele
LOCATION: 97099..97145
OTHER INFORMATION: polymorphic fragment 99-1442-224 SEQ ID71

FEATURE:
NAME/KEY: allele
LOCATION: 97130..97177
OTHER INFORMATION: polymorphic fragment 5-129-144 SEQ ID33

FEATURE:
NAME/KEY: allele
LOCATION: 97130..97177
OTHER INFORMATION: polymorphic fragment 5-129-144 SEQ ID54

FEATURE:
NAME/KEY: allele
LOCATION: 99075..99121
OTHER INFORMATION: polymorphic fragment 5-130-257 SEQ ID55

FEATURE:
NAME/KEY: allele
LOCATION: 99075..99121
OTHER INFORMATION: polymorphic fragment 5-130-257 SEQ ID34

FEATURE:
NAME/KEY: allele
LOCATION: 99094..99140
OTHER INFORMATION: polymorphic fragment 5-130-276 SEQ ID35

FEATURE:
NAME/KEY: allele
LOCATION: 99094..99140
OTHER INFORMATION: polymorphic fragment 5-130-276 SEQ ID56

FEATURE:
NAME/KEY: allele
LOCATION: 103783..103828
OTHER INFORMATION: polymorphic fragment 5-131-395 SEQ ID36

FEATURE:
NAME/KEY: allele
LOCATION: 103783..103828
OTHER INFORMATION: polymorphic fragment 5-131-395 SEQ ID57

FEATURE:
NAME/KEY: allele
LOCATION: 106918..106966
OTHER INFORMATION: polymorphic fragment 5-133-375 SEQ ID37

FEATURE:
NAME/KEY: allele
LOCATION: 106918..106966
OTHER INFORMATION: polymorphic fragment 5-133-375 SEQ ID58

FEATURE:
NAME/KEY: allele
LOCATION: 108084..108130
OTHER INFORMATION: polymorphic fragment 5-135-155 SEQ ID38

FEATURE:
NAME/KEY: allele
LOCATION: 108084..108130
OTHER INFORMATION: polymorphic fragment 5-135-155 SEQ ID59

FEATURE:
NAME/KEY: allele
LOCATION: 108127..108177

```
;   OTHER INFORMATION: polymorphic fragment 5-135-198 SEQ ID39
;   FEATURE:
;   NAME/KEY: allele
;   LOCATION: 108127..108177
;   OTHER INFORMATION: polymorphic fragment 5-135-198 SEQ ID60
;   FEATURE:

Query Match    24.3%;   Score 243;   DB 4;   Length 162450;
Best Local Similarity   82.3%;   Pred. No. 4e-57;
Matches  279;   Conservative   0;   Mismatches  60;   Indels   0;   Gaps   0;

Qy     469 GGCTGGGCACGGTGGCTCATGCCTGCAATCCCAGCACACCCTTGGGAGGCTGAGGCAGGCAGA 528
           |||| |||||||||||||||  ||||||||||||| ||||||||||| ||||||||||  ||||
Db  118499 GCCAGGCACGGGTAGCTCATGCCCATAATCCCACGCACTTTGGGAGGCCGAGGCGGGCAGA 118558

Qy     529 TGACTTGAGGCCAGGAGTTCGAGACCAGCCTGGCCAACATGGTGAAACCCCATCATTTCTACT 588
           | ||| |||| ||||||||||| |||||||||||| ||||  |||||||| |||| ||||||
Db  118559 TCACCTGAGGTCAGGAGTTCGAAACCAGCCTGGGCAACACGGTGAAACCTCATCTCTACT 118618

Qy     589 AAAAATACAAAAAACTAACCAGGCGTGGTGGCGTGGGCTGTAATCCCAGCTACTCGGGAG 648
           |||||  |||||  ||  ||||| ||||||||| ||| || ||||||||||||||  ||||
Db  118619 AAAAATACAAAAAATTAGCCAGACGCATGGTGGCCTTGCTGTGTAATCCCAGCTACTTGGGAG 118678

Qy     649 GCTGAGGCAGGAGAATCCGCTTGAACCGGGGGAGGCGGAGGTTGCAGTGACCTGAGATCACA 708
           ||||||||||| ||||| || | ||||||||||||| ||||||||| ||| ||||||||||
Db  118679 GCTGAGGGCATGAGAATGACTGATCCAGGGCACAGAGGTTGCAGTGAGCTGAACCAAGATCACA 118738

Qy     709 CCACTGCACTCCAGCCTGGGCGACAGAGCGAGACTCCAGCTTAAAAAAAAAAAAAAAAA 768
           ||||| ||  |||| |||||||||||||| |||||| ||||||||||||||||||||||
Db  118739 CCATTGCACTCCAGCCTGGGTGACAGAGCAAGACTCCATCTCAAAAAAAAAAAAAAAACAAA 118798

Qy     769 AAAAAAGGAGACCATCACTGCGTGTCCTGCGCATTCTTACA 807
           ||||||  ||||   ||| |||| || | |||||| || |||
Db  118799 CAAAAACAAAAAATCAATTTTCTTGTTAATTTTTACA 118837
```

```
;   SEQUENCE CHARACTERISTICS:
;   LENGTH: 7676 base pairs
;   TYPE: nucleic acid
;   STRANDEDNESS: double
;   TOPOLOGY: linear
;   MOLECULE TYPE: DNA (genomic)
US-08-451-777A-7

Query Match    24.0%;   Score 240.6;   DB 1;   Length 7676;
Best Local Similarity   84.6%;   Pred. No. 4.6e-57;
Matches  270;   Conservative   0;   Mismatches  49;   Indels   0;   Gaps   0;

Qy     462 AGTCATTGGGCTGGGCACGGTGGCTCATGCCTGCAATCCCAGCACACCTTGGGAGGCTGAGGC 521
           ||  ||| |||||||||||||||||||||||||||||||||||| ||||||||||||||||||
Db    6065 AATGAATGGGGTGGGCGCGGTGGCTCACACCTGTAATCCCAGCACTTTGGGAGGCTGAGGC 6006

Qy     522 AGGCAGGATGACTTGAGGCCAGGAGTTCGAGACCAGCCTGGCCAATAATGGGCGAAACTCAT 581
           ||||||||  ||||||||  ||||||||||||||||||||| ||||||  ||||||||| ||
Db    6005 AGGCAGATTGCCTGAGGTCAGGAGTTCGAGACCAGCCTGGGCCAACATGGCGAAACCCCGT 5946

Qy     582 TTCTACTAAAAATACAAAAAACTAACCAGGCGTGGTGGCTTGTGCCTGTAATCCCAGCTAC 641
           | ||| ||||||||||||||||||| ||||||| |||||| ||| ||| ||||||||||||
Db    5945 CTCTCACTAAAAATACAAAAATTAGCCGGGCGTGGTGGCGCGCGCCTGTAATCCCAGCTAC 5886

Qy     642 TCGGGAGGCTGAGGCAGGAGAATCGCTTGAACCCGGGAGGCGGAGGTTGCAGTGAGCTGA 701
           ||||||||| |||||||| || |||||||||||||||||||| |||||||||||||||||
Db    5885 TCAGGAGGCTGAGGCAGGAGAATTGCTTGAACCTGGGAGGCGGAGGTTGCAGTGAGCTGGA 5826

Qy     702 GATCACACCACTGCACTCCAGCCTGGGCGACAGAGCGAGACTCCAGCTTAAAAAAAAAAA 761
           ||||||||||| |||||||||||||||| ||||||| |||||| |||||||||||||||
Db    5825 GATCATGCCACTGCACTTCAGCCTGGGTGACAGAGTGAGACTCCGTCTCAAAAAAAAAAA 5766

Qy     762 AAAAAAAAAAAAGGAGA 780
           ||  || |||||||||
Db    5765 AGAAGAAAAGAAAGAAAGA 5747
```

```
;  INFORMATION FOR SEQ ID NO: 7:
;  SEQUENCE CHARACTERISTICS:
;   LENGTH: 7676 base pairs
;   TYPE: nucleic acid
;   STRANDEDNESS: double
;   TOPOLOGY: linear
;  MOLECULE TYPE: DNA (genomic)
US-08-451-778A-7

Query Match       24.0%;  Score 240.6;  DB 2;  Length 7676;
Best Local Similarity   84.6%;  Pred. No. 4.6e-57;
Matches  270;  Conservative   0;  Mismatches  49;  Indels   0;  Gaps   0;

Qy    462 AGTCATTGGCTGGGCACGGTGGCCTCATGCCTGCAATCCCAGCACCTTGGGAGGCTGAGGC 521
          | | ||| ||||| ||||| ||||||||| || |||| ||||||||| |||||| |||||||
Db   6065 AATGAATGGGTGGGGCGGGTGGCTCACACCTGTAATCCCAGCACTTTGGGAAGCTGAGGC 6006

Qy    522 AGGCAGAATGACTTGAGGCCAGGAGTTCGAGACCAGCCTGGCCAATATGGCGAAACCTCAT 581
          ||||||||| | |||||| |||||| | |||||||||||| ||| |||||||||||  | |
Db   6005 AGGCAGAATTGCCTGAGGTCAGGAGTCAGGACCAGCCTGAGACCAAACATGGCGAAACCCCGT 5946

Qy    582 TTCTACTAAAAATACAAAAAACTAACCAGGCGGTGGTGGCGTGGCCTGTAGTCCCAGCTAC 641
          ||||||||||||||||||| | ||||| ||| |||| ||||| |||||||||||||||||
Db   5945 CTCTACTAAAAATACAAAAAGTTAGCCGGGCATCGTGGCGCGCCTGTAGTCCCAGCTAC 5886

Qy    642 TCGGGAGGCTGAGGCAGGAGAATCGGTTGAACCGGGAGGCGGAGGTTGCAGTGAGCTGA 701
          || ||||||||||||||||||||| | ||||||| |||||||||||||||||||| ||
Db   5885 TCAGGAGGCTGAGGCAGGAGAATTGCTTGAACCTGGGAGGCGGAGGTTGCAGTGAGTGGA 5826

Qy    702 GATCACACCACTGCACTCCAGCCTGGGCGACAGAGCGAGACTCCAGCTTAAAAAAAAAA 761
          ||||| |||||||||| || ||||||| |||||| |||| ||| | |||||||||||||
Db   5825 GATCATGCCACTGCACTTCACCCTGGGTGACAGAGTGAGGGTCTCCCTCTCAAAAAAAAAAA 5766

Qy    762 AAAAAAAAAAAAAAAGGAGA 780
          |||| ||| ||| |||
Db   5765 AGAAAGAAAGAAAGAAAGA 5747
```

```
RESULT 4
US-08-998-208-7/c
Sequence 7, Application US/0898208
Patent No. 5880105
GENERAL INFORMATION:
 APPLICANT: Bergsma, Derk J.
 APPLICANT: Stambolian, Dwight
 TITLE OF INVENTION: Human Galactokinase Gene
 NUMBER OF SEQUENCES: 33
 CORRESPONDENCE ADDRESS:
   ADDRESSEE: Smithkline Beecham Corp./Corporate
              Intellectual Property
   STREET: 709 Swedeland Road/UW2220
   CITY: King of Prussia
   STATE: Pennsylvania
   COUNTRY: USA
   ZIP: 19406-0939
 COMPUTER READABLE FORM:
   MEDIUM TYPE: Floppy disk
   COMPUTER: IBM PC compatible
   OPERATING SYSTEM: PC-DOS/MS-DOS
   SOFTWARE: Patentin Release #1.0, Version #1.30
 CURRENT APPLICATION DATA:
   APPLICATION NUMBER: US/08/998,208
   FILING DATE:
   CLASSIFICATION: 435
 PRIOR APPLICATION DATA:
   APPLICATION NUMBER: 08/451,777
   FILING DATE: 26-MAY-1995
   APPLICATION NUMBER: PCT/US94/10825
   FILING DATE: 23-SEP-1994
 ATTORNEY/AGENT INFORMATION:
   NAME: Eagle, Alissa M.
   REGISTRATION NUMBER: 37,126
   REFERENCE/DOCKET NUMBER: P50268-1B
```

```
; TELECOMMUNICATION INFORMATION:
;   TELEPHONE: 610-270-5364
;   TELEFAX: 610-270-5090
; INFORMATION FOR SEQ ID NO: 7:
; SEQUENCE CHARACTERISTICS:
;   LENGTH: 7676 base pairs
;   TYPE: nucleic acid
;   STRANDEDNESS: double
;   TOPOLOGY: linear
;  MOLECULE TYPE: DNA (genomic)
US-08-998-208-7

Query Match       24.0%;  Score 240.6;  DB 2;  Length 7676;
Best Local Similarity   84.6%;  Pred. No. 4.6e-57;
Matches  270;  Conservative   0;  Mismatches  49;  Indels   0;  Gaps   0;

Qy    462 AGTCATTGGCTGGGCACGGTGGCCTCATGCCTGCAATCCCAGCACCTTGGGAGGCTGAGGC 521
          | | ||| ||||| ||||| ||||||||| || |||| ||||||||| |||||| |||||||
Db   6065 AATGAATGGGTGGGGCGGGTGGCTCACACCTGTAATCCCAGCACTTTGGGAAGCTGAGGC 6006

Qy    522 AGGCAGAATGACTTGAGGCCAGGAGTTCGAGACCAGCCTGGCCAATATGGCGAAACCTCAT 581
          ||||||||| | |||||| |||||| | |||||||||||| ||| |||||||||||  | |
Db   6005 AGGCAGAATTGCCTGAGGTCAGGAGTCAGGACCAGCCTGAGACCAAACATGGCGAAACCCCGT 5946

Qy    582 TTCTACTAAAAATACAAAAAACTAACCAGGCGGTGGTGGCGTGGCCTGTAGTCCCAGCTAC 641
          ||||||||||||||||||| | ||||| ||| |||| ||||| |||||||||||||||||
Db   5945 CTCTACTAAAAATACAAAAAGTTAGCCGGGCATCGTGGCGCGCCTGTAGTCCCAGCTAC 5886

Qy    642 TCGGGAGGCTGAGGCAGGAGAATCGGTTGAACCGGGAGGCGGAGGTTGCAGTGAGCTGA 701
          || ||||||||||||||||||||| | ||||||| |||||||||||||||||||| ||
Db   5885 TCAGGAGGCTGAGGCAGGAGAATTGCTTGAACCTGGGAGGCGGAGGTTGCAGTGAGTGGA 5826

Qy    702 GATCACACCACTGCACTCCAGCCTGGGCGACAGAGCGAGACTCCAGCTTAAAAAAAAAA 761
          ||||| |||||||||| || ||||||| |||||| |||| ||| | |||||||||||||
Db   5825 GATCATGCCACTGCACTTCACCCTGGGTGACAGAGTGAGGGTCTCCCTCTCAAAAAAAAAAA 5766

Qy    762 AAAAAAAAAAAAAAAGGAGA 780
          |||| ||| ||| |||
Db   5765 AGAAAGAAAGAAAGAAAGA 5747
```

```
RESULT 5
PCT-US95-06743-7/c
Sequence 7, Application PC/TUS906743
GENERAL INFORMATION:
 APPLICANT: Bergsma, Derk J.
 APPLICANT: Stambolian, Dwight
 TITLE OF INVENTION: Human Galactokinase Gene
 NUMBER OF SEQUENCES: 32
 CORRESPONDENCE ADDRESS:
   ADDRESSEE: SmithKline Beecham Corp./Corporate
              Intellectual Property
   STREET: 709 Swedeland Road/UW2220
   CITY: King of Prussia
   STATE: Pennsylvania
   COUNTRY: USA
   ZIP: 19406-0939
 COMPUTER READABLE FORM:
   MEDIUM TYPE: Floppy disk
   COMPUTER: IBM PC compatible
   OPERATING SYSTEM: PC-DOS/MS-DOS
   SOFTWARE: Patentin Release #1.0, Version #1.30
 CURRENT APPLICATION DATA:
   APPLICATION NUMBER: PCT/US95/06743
   FILING DATE:
   CLASSIFICATION:
 PRIOR APPLICATION DATA:
   APPLICATION NUMBER: PCT/US94/10825
   FILING DATE: 23-SEP-1994
 ATTORNEY/AGENT INFORMATION:
   NAME: Sutton, Jeffrey A.
   REGISTRATION NUMBER: 34,028
   REFERENCE/DOCKET NUMBER: P50268-1
```

```
;   TELECOMMUNICATION INFORMATION:
;       TELEPHONE: 610-270-5024
;       TELEFAX: 610-270-5090
;   INFORMATION FOR SEQ ID NO: 7:
;       SEQUENCE CHARACTERISTICS:
;           LENGTH: 7676 base pairs
;           TYPE: nucleic acid
;           STRANDEDNESS: double
;           TOPOLOGY: linear
;       MOLECULE TYPE: DNA (genomic)
;   PCT-US95-06743-7

Query Match             84.6%;  Score 240.6;  DB 5;  Length 7676;
Best Local Similarity   84.6%;  Pred. No. 4.6e-57;
Matches 270;  Conservative  0;  Mismatches 49;  Indels  0;  Gaps  0;

Qy    462  AGTCATTGGCTGGGCACCGTGGCCTCAATCCCAGCACCTTGGGAGGCTGAGGC  521
           |||  | ||||||  ||||||||||||||||||||||||||||||| | |||||
Db   6065  AATGAATGGGTGGGCGCCGGTGGCTCACACCTGTAATCCCAGCACTTTGGGAGGCTGAGGC  6006

Qy    522  AGGCAGAGATGACTTGAGGCCAGGAGAGTTCGAGACCAGCACCCAATATGGGGAAACCTCAT  581
           |||||||| | || ||||||||||||||||||| | |||||| |||||||| ||||||||||
Db   6005  AGGCAGAGATTGCCTGAGGTCAGGAGGACCTCGAGACCAGCCTGGCCAACATGGCGAAACCCCGT  5946

Qy    582  TTCTACTAAAAAATACAAAAAACTAACCAGGCGTGGTGGCGTTGCCTGTGTTGTAATCCCAGCTAC  641
           |||||||||   ||| |||||||||| |||||||||||  | ||||||||||||||||||||||||
Db   5945  CTCTACTAAAAAATACAAAAAGTTAGCCGGGCGTGGCCAGGCGCCTGTAATCCCAGCTAC  5886

Qy    642  TCGGGAGGCTGAGGCAGGAGAATCGCTTGAACCCGGGAGGCGGAGGTTGCAGTGAGCTGA  701
           |||||||||||||||||||| |||||||||||||||||||||||||||||  ||||||  ||
Db   5885  TCAGGAGGCTGAGGCAGGAGAATTGCTTGAACCTGGGAGGCGGAGGTTGCAGTGAGTGGA  5826

Qy    702  GATCACCACCACTGCACTCCAGCCTGGGCGACAGAGCGAGACTCCAGCTTAAAAAAAAAA  761
           |||| | ||| |||||| ||||||||||||| ||||| ||| ||| | ||||||||||||
Db   5825  GATCATGCCACTGCACTTCAGCCTGGGTGACAGAGTGAGACTCCGTCTCAAAAAAAAAAAA  5766

Qy    762  AAAAAAAAAAAAAAAGGAGA  780
           |||||||   ||| |||
Db   5765  AGAAAGAAAGAAAGAAAGA  5747
```

```
;   RESULT 6
;   US-09-345-882-1/c
;   Sequence 1, Application US/09345882
;   Patent No. 6399373
;   GENERAL INFORMATION:
;       APPLICANT: Bougueleret, Lydie
;       TITLE OF INVENTION: A NUCLEIC ACID ENCODING A RETINOBLASTOMA BINDING PROTEIN (RBP-7)
;       TITLE OF INVENTION: AND POLYMORPHIC MARKERS ASSOCIATED WITH SAID NUCLEIC ACID.
;       FILE REFERENCE: GENSET.031A
;       CURRENT APPLICATION NUMBER: US/09/345,882
;       CURRENT FILING DATE: 1999-06-30
;       PRIOR APPLICATION NUMBER: US 60/091,315
;       PRIOR FILING DATE: 1998-06-30
;       PRIOR APPLICATION NUMBER: US 60/111,909
;       PRIOR FILING DATE: 1998-12-10
;       NUMBER OF SEQ ID NOS: 140
;       SOFTWARE: Patent.pm
;   SEQ ID NO 1
;       LENGTH: 162450
;       TYPE: DNA
;       ORGANISM: Homo sapiens
;       FEATURE:
;           NAME/KEY: allele
;           LOCATION: 72794
;       OTHER INFORMATION: 5-124-273   :  polymorphic  base  A  or  G
;       FEATURE:
;           NAME/KEY: allele
;           LOCATION: 88073
;       OTHER INFORMATION: 5-127-261   :  polymorphic  base  A  or  C
;       FEATURE:
;           NAME/KEY: allele
;           LOCATION: 90842
```

```
;       OTHER INFORMATION: 99-1437-325   :  polymorphic  base  A  or  G
;       FEATURE:
;           NAME/KEY: allele
;           LOCATION: 93714
;       OTHER INFORMATION: 5-128-60   :  polymorphic  base  deletion  of  GT
;       FEATURE:
;           NAME/KEY: allele
;           LOCATION: 97122
;       OTHER INFORMATION: 99-1442-224   :  polymorphic  base  G  or  T
;       FEATURE:
;           NAME/KEY: allele
;           LOCATION: 9752
;       OTHER INFORMATION: 5-129-144   :  polymorphic  base  deletion  of  T
;       FEATURE:
;           NAME/KEY: allele
;           LOCATION: 99098
;       OTHER INFORMATION: 5-130-257   :  polymorphic  base  A  or  G
;       FEATURE:
;           NAME/KEY: allele
;           LOCATION: 99117
;       OTHER INFORMATION: 5-130-276   :  polymorphic  base  A  or  G
;       FEATURE:
;           NAME/KEY: allele
;           LOCATION: 103806
;       OTHER INFORMATION: 5-131-395   :  polymorphic  base  A  or  T
;       FEATURE:
;           NAME/KEY: allele
;           LOCATION: 106940
;       OTHER INFORMATION: 5-133-375   :  polymorphic  base  A
;       FEATURE:
;           NAME/KEY: allele
;           LOCATION: 108106
;       OTHER INFORMATION: 5-135-155   :  polymorphic  base  insertion  of  A
;       FEATURE:
;           NAME/KEY: allele
;           LOCATION: 108149
;       OTHER INFORMATION: 5-135-198   :  polymorphic  base  insertion  of  GTTTT
;       FEATURE:
;           NAME/KEY: allele
;           LOCATION: 108308
;       OTHER INFORMATION: 5-135-357   :  polymorphic  base  A  or  G
;       FEATURE:
;           NAME/KEY: allele
;           LOCATION: 108471
;       OTHER INFORMATION: 5-136-174   :  polymorphic  base  C  or  T
;       FEATURE:
;           NAME/KEY: allele
;           LOCATION: 134134
;       OTHER INFORMATION: 5-140-120   :  polymorphic  base  C  or  T
;       FEATURE:
;           NAME/KEY: allele
;           LOCATION: 134362
;       OTHER INFORMATION: 5-140-348   :  polymorphic  base  insertion  of  A
;       FEATURE:
;           NAME/KEY: allele
;           LOCATION: 134374
;       OTHER INFORMATION: 5-140-361   :  polymorphic  base  insertion  of  CA
;       FEATURE:
;           NAME/KEY: allele
;           LOCATION: 143328
;       OTHER INFORMATION: 5-143-84   :  polymorphic  base  A  or  G
;       FEATURE:
;           NAME/KEY: allele
;           LOCATION: 143345
;       OTHER INFORMATION: 5-143-101   :  polymorphic  base  A  or  G
;       FEATURE:
;           NAME/KEY: allele
;           LOCATION: 150329
;       OTHER INFORMATION: 5-145-24   :  polymorphic  base  A  or  G
;       FEATURE:
;           NAME/KEY: allele
;           LOCATION: 160031
;       OTHER INFORMATION: 5-148-352   :  polymorphic  base  G  or  T
```

FEATURE:
NAME/KEY: allele
LOCATION: 72771..72817
OTHER INFORMATION: polymorphic fragment 5-124-273 SEQ ID30
FEATURE:
NAME/KEY: allele
LOCATION: 72771..72817
OTHER INFORMATION: polymorphic fragment 5-124-273 SEQ ID51
FEATURE:
NAME/KEY: allele
LOCATION: 88050..88096
OTHER INFORMATION: polymorphic fragment 5-127-261 SEQ ID31
FEATURE:
NAME/KEY: allele
LOCATION: 88050..88096
OTHER INFORMATION: polymorphic fragment 5-127-261 SEQ ID52
FEATURE:
NAME/KEY: allele
LOCATION: 90819..90865
OTHER INFORMATION: complement polymorphic fragment 99-1437-325 SEQ ID70
FEATURE:
NAME/KEY: allele
LOCATION: 90819..90865
OTHER INFORMATION: complement polymorphic fragment 99-1437-325 SEQ ID49
FEATURE:
NAME/KEY: allele
LOCATION: 93690..93736
OTHER INFORMATION: polymorphic fragment 5-128-60 SEQ ID53
FEATURE:
NAME/KEY: allele
LOCATION: 93690..93736
OTHER INFORMATION: polymorphic fragment 5-128-60 SEQ ID32
FEATURE:
NAME/KEY: allele
LOCATION: 97099..97145
OTHER INFORMATION: polymorphic fragment 99-1442-224 SEQ ID50
FEATURE:
NAME/KEY: allele
LOCATION: 97099..97145
OTHER INFORMATION: polymorphic fragment 99-1442-224 SEQ ID71
FEATURE:
NAME/KEY: allele
LOCATION: 97130..97177
OTHER INFORMATION: polymorphic fragment 5-129-144 SEQ ID33
FEATURE:
NAME/KEY: allele
LOCATION: 97130..97177
OTHER INFORMATION: polymorphic fragment 5-129-144 SEQ ID54
FEATURE:
NAME/KEY: allele
LOCATION: 99075..99121
OTHER INFORMATION: polymorphic fragment 5-130-257 SEQ ID34
FEATURE:
NAME/KEY: allele
LOCATION: 99075..99121
OTHER INFORMATION: polymorphic fragment 5-130-257 SEQ ID55
FEATURE:
NAME/KEY: allele
LOCATION: 99094..99140
OTHER INFORMATION: polymorphic fragment 5-130-276 SEQ ID35
FEATURE:
NAME/KEY: allele
LOCATION: 99094..99140
OTHER INFORMATION: polymorphic fragment 5-130-276 SEQ ID56
FEATURE:
NAME/KEY: allele
LOCATION: 103783..103828
OTHER INFORMATION: polymorphic fragment 5-131-395 SEQ ID36
FEATURE:
NAME/KEY: allele
LOCATION: 103783..103828
OTHER INFORMATION: polymorphic fragment 5-131-395 SEQ ID57
FEATURE:

NAME/KEY: allele
LOCATION: 106918..106966
OTHER INFORMATION: polymorphic fragment 5-133-375 SEQ ID37
FEATURE:
NAME/KEY: allele
LOCATION: 106918..106966
OTHER INFORMATION: polymorphic fragment 5-133-375 SEQ ID58
FEATURE:
NAME/KEY: allele
LOCATION: 108084..108130
OTHER INFORMATION: polymorphic fragment 5-135-155 SEQ ID38
FEATURE:
NAME/KEY: allele
LOCATION: 108084..108130
OTHER INFORMATION: polymorphic fragment 5-135-155 SEQ ID59
FEATURE:
NAME/KEY: allele
LOCATION: 108127..108177
OTHER INFORMATION: polymorphic fragment 5-135-198 SEQ ID39
FEATURE:
NAME/KEY: allele
LOCATION: 108127..108177
OTHER INFORMATION: polymorphic fragment 5-135-198 SEQ ID60
FEATURE:

Query Match   23.9%; Score 238.8; DB 4; Length 162450;
Best Local Similarity   86.3%; Pred. No. 5.8e-56;
Matches 264; Conservative   0; Mismatches   42; Indels   0; Gaps   0;

```
Qy        467 TTGGCTGGGCCACGGTGGCTCATGCCTGCAAATCCCAGCACCTTGGGAGGCTGAGGCAGGCA 526
              |  ||| |||| |||| |||||||  | | ||| |||||||||| | | |||||||||||||
Db      72459 TGGCGTGAGCACACGTGGCTCTCACGCCTGCAATCCCAGCACTTTGAGGGGCTGAGGCAGGTG 72400

Qy        527 GATGACTTGAGGCGCCAGGAGTTCGAGACACAGCCTGGCCAATATGGCGAAACCTCATTTCTA 586
              ||| |||||||| |||||||||||||||||| |||||||||||| || | |||||| |||| ||
Db      72399 GATCACTTGAGGTCAGGAGTTCGAGACCCAGCCTGGCCAACATGGTGAAACCCCGTCTCTTA 72340

Qy        587 CTAAAAATACAAAAAACTAACCAGGCGTGGTGGTGGGCTTGTGCCTGTAATCCCAGCTACTCGGG 646
              | |||| |||||||  || | || || ||| |||| |||||||  ||||||||||||||||||||
Db      72339 CCAAAAGTACAAAAATTGGGCAGCAGGCGCGGCGCCTGTGGTCCCAGCTACTCGGG 72280

Qy        647 AGGCTGAGGCAGGAGAATCGCTTGAACCCGGGAGGCGGAGGTTGCAGTGAGCTGAGATCA 706
              |||| ||||| ||||||| | ||||| || ||||||| |||||||||||||||| |||||
Db      72279 AGGGTGAGGCCAGGAGAAATTGGCTTGAGCCTGGGAGGTTGCAGTGAGCCCAGATCG 72220

Qy        707 CACCACTGCACTCCAGCCTGGGCGACAGAGCGAGACTCCAGCTTAAAAAAAAAAAAAAA 766
              | ||||||| |||||||||||| |||||||| ||| || |||| |||||||||||||||
Db      72219 CGCCACTGCCACTCCAGCCTGGGTGACAGAGCAAGAGACTCTGCCTCAAAAAAAAAAAAA 72160

Qy        767 AAAAAA 772
              ||||||
Db      72159 AAAAAA 72154
```

```
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
  APPLICATION NUMBER: US/08/969,125B
  FILING DATE: 12-No.
  CLASSIFICATION: <Unknown>
PRIOR APPLICATION DATA:
  APPLICATION NUMBER: 6143871-1997
  FILING DATE: 13-DEC-1996
  APPLICATION NUMBER: GB 9625899.1
  OTHER APPLICATION DATA:
ATTORNEY/AGENT INFORMATION:
  NAME: WILSON, MARY J.
  REGISTRATION NUMBER: 32,955
  REFERENCE/DOCKET NUMBER: 1430-179
TELECOMMUNICATION INFORMATION:
  TELEPHONE: (703) 816-4000
  TELEFAX: (703) 816-4100
INFORMATION FOR SEQ ID NO: 8
SEQUENCE CHARACTERISTICS:
  LENGTH: 4038 base pairs
  TYPE: nucleic acid
  STRANDEDNESS: single
  TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
FRAGMENT TYPE: linear
FEATURE:
  NAME/KEY: CDS
  LOCATION: 43..1323
SEQUENCE DESCRIPTION: SEQ ID NO: 8
US-08-969-125-8
```

```
Query Match        23.7%;   Score 237.2;  DB 3;   Length 4038;
Best Local Similarity  83.5%;   Pred. No. 3e-56;
Matches  269;  Conservative   0;  Mismatches  53;  Indels   0;  Gaps   0;

Qy   467 TTGGCTGGGCACGGTGGCCTCATGCCTTGCCAATCCCAGCAACCTTGGGAGGCCGAGGCTG 517
         ||||||| ||||||||||||||| ||||| ||||||||||||||||||||||| |||||||
Db  1834 TCGGCTGGGGTCCGGTGGCTCACGCCTGTAATCCCAGCAATTTGGGAGGTCCGGAGGGCGG 1893

Qy   527 GATGACTTGAGGCGCAGGAGAGTTCAGGACCAGCAGCCTGGCCAATATGGCGAAACCTCATTTCTA 586
         || ||| || |||||||| |||||||||||||||| ||| ||| |||||  || |  ||| ||
Db  1894 GATCACTCGAGGTCAGGAGTTCGAGACCAGCCTGACCAACAAAATGGTGAAACCTCCTCTCTA 1953

Qy   587 CTAAAAATACAAAAAATTACCAGGCGTGGTGGCTTGTGCCTGTAATCCCAGCTACTCGGG 646
         |||||||||||||||||| |||  ||||||||| |||||||||||||||||| || ||||
Db  1954 CTAAAAACTACAAAAAATTAACTGGGTGTGGTGGCGCGTGCCTGTAATCCCAGCTACTCGGG 2013

Qy   647 AGGCTGAGGCAGGAGAATCGCTTGAACCGGGGAGGCGGAGGTTGCAGTGAGCTGAGATCA 706
         | ||||||| ||||| |||  | |||| ||||||| |||||| |||||||||||| |||
Db  2014 AAGCTGAGGCAGGTGAATTGTTTGAACCTGGGAGGTTGCAGTGAGCAGAGATCA 2073

Qy   707 CACCACTGCACTCCAGCCTGGGCGACAGCAGCGAGACTCCAGCTTAAAAAAAAAAAAAAA 766
         |||||||| ||||||||||||| ||||| |||||||||||| |||||||||||||||
Db  2074 CACCACTGCACTCTAGCCTGGGTGACAGAGCAAGACTCTGTCTCAAAAAAACAAAAACAAAAAC 2133

Qy   767 AAAAAAAAAAAGGAGACCATCACT 788
         ||||||||||||| | | | |||
Db  2134 AAAACAAAACAAAAAACCTCT 2155
```

```
RESULT 8
US-09-800-960-3/c
Sequence 3, Application US/09800960
Patent No. 6387677
GENERAL INFORMATION:
APPLICANT: YE, Jane et al.
TITLE OF INVENTION: ISOLATED HUMAN KINASE PROTEINS, NUCLEIC
TITLE OF INVENTION: ACID MOLECULES ENCODING HUMAN KINASE PROTEINS, AND USES
TITLE OF INVENTION: THEREOF
FILE REFERENCE: CL001158
CURRENT APPLICATION NUMBER: US/09/800,960
CURRENT FILING DATE: 2001-03-08
NUMBER OF SEQ ID NOS: 4
SOFTWARE: FastSEQ for Windows Version 4.0
```

```
SEQ ID NO 3
  LENGTH: 62804
  TYPE: DNA
  ORGANISM: Human
FEATURE:
  NAME/KEY: misc_feature
  LOCATION: (1)..(62804)
  OTHER INFORMATION: n = A,T,C or G
US-09-800-960-3
```

```
Query Match        23.7%;   Score 237.2;  DB 4;   Length 62804;
Best Local Similarity  82.4%;   Pred. No. 1.1e-55;
Matches  272;  Conservative   0;  Mismatches  58;  Indels   0;  Gaps   0;

Qy   458 AGACAGTCATTGGCTCGGGCACGGTGGCCTCATGCCTGCAATCCCAGCAACCTTGGGAGGCTG 517
         ||| || ||||||||||||||||||||||||| |||| ||| |||||||||||||||||||||
Db 48961 ACATGGTGGGTGGGTGGGCACGGTGGCTCACACCTGTAATCCCAGCACTTTGGGAGGCCA 48902

Qy   518 AGGCAGGCAGAGATGACTTGAGGCCAGGAGTTCGAGACCAGCCAGCCTGGCCAATATGGCGAAACC 577
         |||||| ||  ||||||  ||| |||||||||||||||||| || ||||||||  |||| ||| ||
Db 48901 AGGCGGGTGGATCACCTGAGGTCAGGAGTTCAAGACCAGCCTGGCCAACATGGTGAAACC 48842

Qy   578 TCATTTCTACTAAAAATACAAAAAACTAACCAGGCGTGGTGGCTTGTTGCCTGTAATCCCAG 637
         ||||| | |||||||| ||||||  |||||||||||||||||||| || |||||||||||||
Db 48841 CCGTCTCTACTAAAAATACAAAAAATTAGCCGGGCATGGTGGCACATGCCTGTAATCCCAG 48782

Qy   638 CTACTCGGGAGGCTGAGGCAGGAGAATCGCTTGAACCGGGGAGGCGGAGGTTGCAGTGAG 697
         ||||||||||||||||||| ||| ||||| ||||||||||| |||| ||||||| |||||
Db 48781 CTACTCGGGAGGCTGAGACGAGAATTGCTTGAACCTGGGAGGCGGAGGTTGCAGTGAG 48722

Qy   698 CTGAGATCACACCACTGCACTCCAGCCTGGGCGACAGAGCGAGACTCCAGCTTAAAAAAA 757
         |||| ||||||||||||| |||||||||||| |||||||| |||||||||||||||||||
Db 48721 CCAAGATTGCACCACTGCACTCCAGCCTGGGCAACAGAGCAAGACTCCATCTCAAAAAAC 48662

Qy   758 AAAAAAAAAAAAAAAAAAAAAAAAGGAGACCATCAC 787
         |||   || |||| |||||  ||| |||| |||||
Db 48661 CAAAAGCAAACGAACAAAACAAAACATTAC 48632
```

```
RESULT 9
US-09-729-995-3
Sequence 3, Application US/09729995
Patent No. 6426206
GENERAL INFORMATION:
APPLICANT: WEI, Ming-Hui et al
TITLE OF INVENTION: ISOLATED HUMAN KINASE PROTEINS, NUCLEIC
TITLE OF INVENTION: ACID MOLECULES ENCODING HUMAN KINASE PROTEINS, AND USES
TITLE OF INVENTION: THEREOF
FILE REFERENCE: CL000904
CURRENT APPLICATION NUMBER: US/09/729,995
CURRENT FILING DATE: 2000-12-06
NUMBER OF SEQ ID NOS: 4
SOFTWARE: FastSEQ for Windows Version 4.0
SEQ ID NO 3
  LENGTH: 29629
  TYPE: DNA
  ORGANISM: Human
US-09-729-995-3
```

```
Query Match        23.7%;   Score 237;   DB 4;   Length 29629;
Best Local Similarity  84.2%;   Pred. No. 8.5e-56;
Matches  267;  Conservative   0;  Mismatches  50;  Indels   0;  Gaps   0;

Qy   470 GCTGGGCACGGTGGCCTCATGCCTGCAATCCCAGCACCTTGGGAGGCCGAGGCAGGCAGAT 529
         |||||||||||||| || || |||||||||| ||||| ||||||||| ||||| ||||||
Db 11631 GCTGGGGCGGGTGGCTCACGCCTGTAATCCCAGCACTTTGGGAGGCCGAGGTGGGTGGAT 11690

Qy   530 GACTTGAGGCCAGGAGTTCGAGACCAGCCTGGCCAATATGGCGAAACCTCATTTCTACTA 589
         || ||| ||||||||||||| ||||||||||||| || |||||||||||||||| ||||
Db 11691 CACCTGAGGTCAGGAGTTCAAGACCAGCCTGACCAACATGGCGAAACCCGTGTCTCTACTA 11750

Qy   590 AAAATACAAAAACTAACCAGCGCGTGGTGGTGGCTTGTGCCTGTAATCCCAGCTACTCGGGAGG 649
         |||| |||||||||||||||||||||||||||| |||||||||||||||| ||||||||||||
```

Db 11751 AAAATACAAAAAGTAGCCGGGCGGTGGTGGCGGGCAGAGACACCTGTAATCCCAACTACTCGGGAGG 11810

Qy 650 CTGAGGCAGGAGAATCGCTTGAACCCGGGAGGCAGGAGGTTGCAGTGAGCTGAGATCACAC 709
      ||||| ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 11811 CTGAGGCAGGAGAATCTCTTGAACCTGGAGGTGGAGGCTGCAGTGAGCCGAGATCACGC 11870

Qy 710 CACTGCGACTCCAGCCTGGGCGACAGAGCGAGACTCCAGCTTAAAAAAAAAAAAAAAAAAAA 769
      ||||||||||||||||||||||||||||| |||||||| ||||||||||||||||||||||
Db 11871 CATTGCGACTCCAGCCTGGGGACAGAGTGAGACTCCGTTCAAAAAAAAAAAAAAAAAAAC 11930

Qy 770 AAAAAAGGAGACCATCA 786
      |||| ||||||| || ||
Db 11931 CAAAAAAAAAAAACAGCA 11947

RESULT 10
US-09-851-896-3/c
; Sequence 3, Application US/09851896
; Patent No. 6410325
; GENERAL INFORMATION:
; APPLICANT: C. Frank Bennett
; APPLICANT: Susan M. Freier
; APPLICANT: Andrew T. Watt
; TITLE OF INVENTION: ANTISENSE MODULATION OF PHOSPHOLIPASE A2, GROUP VI (CA2+-INDEPEND
; TITLE OF INVENTION: EXPRESSION
; FILE REFERENCE: RTS-0220
; CURRENT APPLICATION NUMBER: US/09/851,896
; CURRENT FILING DATE: 2001-05-08
; NUMBER OF SEQ ID NOS: 89
; SEQ ID NO 3
; LENGTH: 70000
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:

Query Match   23.7%;   Score 236.8;   DB 4;   Length 70000;
Best Local Similarity   80.5%;   Pred. No. 1.4e-55;
Matches 277;   Conservative 0;   Mismatches 67;   Indels 0;   Gaps 0;

Qy 445 CACCCCTCTCAAGAGACAGTCATTGCCTGGCCACGGTGGGCTCATGCCTGCAATCCCAGCA 504
      ||||||||| ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 12176 CATGCCCCTTAAAGAGCCAGGGCTAGGCCAGGTGGCTCACGCCTGCAATCCCAGCA 12117

Qy 505 CCTTGGGAGGCTGAGGCCAGGAGGATGAACTTGAGGCCAGGAGTTCGAGAACCAGCCTG 564
      ||||| ||||| |||||||||||||||||||||||||||||||||| ||||||||||||
Db 12116 CTTTGGGAGGCCAAGGTGGGCGGATCACCTGAGGTCAGGAGTTCGAGACCAGCCTG 12057

Qy 565 ATATGCGAAACCTCATTTCTACTAAAAATACAAAAAACTAACCACGCGTGGTGGTGCTTGTG 624
      ||||| ||||||||||||||||||||||||||||||||||||||||||||||| ||
Db 12056 ATATGATGAAAACCCCGTCTCTTCGCTAAAAATACAAAAATCAGCTGGGCCGTGGTGGTG 11997

Qy 625 CCTGTAATCCCAGCTACTCGGGAGGCTGAGGCAGGAGAATCGCTTGAACCCGGGAGGCAG 684
      |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 11996 CCTGTAATCCCAGCTACTCAGGAGGCTGAGGCAGGAGAATGGCTTGAACCCTGGGAGGCAG 11937

Qy 685 AGGTTGCAGTGAGCTGAGATCACACCACTCCAGCCTGGGCGACAGAGCGAGACTC 744
      |||||||||||||||||||||||| |||||||||||||||||||||||||||| |||||
Db 11936 AGGTTGCAGTGAGCCCGAGATTGCGCCACTGCACTCCAGCCTGGGCAACAGAGCAAGACTG 11877

Qy 745 CAGCTTAAAAAAAAAAAAAAAAAAAAAAAGGAGACCATCACT 788
      | |||||| |||| |||||| | |||||| | ||
Db 11876 TCTCAAAAAAAACAAAAAACCAAAAATAAAGAGCCAGGGGCT 11833

RESULT 11
US-09-078-294-9/c
; Sequence 9, Application US/09078294
; Patent No. 6265211
; GENERAL INFORMATION:
; APPLICANT: Choo, Kong-Hong Andy
; APPLICANT: Du Sart, Desiree
; APPLICANT: Cancilla, Michael R.

; TITLE OF INVENTION: A NOVEL NUCLEIC ACID MOLECULE
; FILE REFERENCE: Davies Col
; CURRENT APPLICATION NUMBER: US/09/078,294
; CURRENT FILING DATE: 1998-05-13
; NUMBER OF SEQ ID NOS: 29
; SOFTWARE: Patentin Ver. 2.0
; SEQ ID NO 9
; LENGTH: 1701
; TYPE: DNA
; ORGANISM: BAC-F2 contig 5
US-09-078-294-9

Query Match   23.6%;   Score 236.4;   DB 4;   Length 1701;
Best Local Similarity   82.9%;   Pred. No. 3.4e-56;
Matches 282;   Conservative 0;   Mismatches 56;   Indels 2;   Gaps 1;

Qy 441 CTTACACCCCTCTCAAGAGACAGTCATTGGCTGGGCACGGTGGGCTCATGCCTGCAATCC 500
      |||| ||||||||||||||| |||||||||| ||||| ||||||||||||||||||||||
Db 1428 CTTATTTCTCTGTTTAAAAGAAGAGTAGGCCGGGCATGGTGGCGCACACCTGTAATCC 1369

Qy 501 AGCACCTTGGGAGGCTGAGGCAGGCAGGAGATGACTTGAGGCCAGGAGTTCGAGAACCAGCCTG 560
      ||||||||||||||||||||||||||||||| || |||||||||||||||||||||||||
Db 1368 AGCACTTGGGAGGCTGAGGCAGGCAGGGGATCAC--AAGGTCAGGAGTTTGACACCAGCCTG 1311

Qy 561 GCCAATATGGCGAAACCTCATTTCTACTAAAAATACAAAAACTAACCAGGCGTGGTGGCT 620
      |||| || ||||||| |||||||||||||||||||||||||||||||||||||||||
Db 1310 ACCAACATGGTGAAAACCCCATCTCTACTAAAAATTACAAAAATTAGCCGGGCGTGGTGGCG 1251

Qy 621 TGTGCCTGTAATCCCAGCTACTCGGGAGGCTGAGGCAGGAGAATCGCTTGAACCGGGAG 680
      ||| || || |||||||||||| ||||||||||||||||||||| ||||||||||||||
Db 1250 CATGCCTGTAATCCCAGCTACTTGGGAGGCTGAGGCAGGAGAATCGCTTGAACCCGGGAG 1191

Qy 681 GCAGAGGTTGCAGTGAGCTGAGATCACACCACCTGCACTCCAGCCTGGGCGACAGAGCGAG 740
      |||||||||||||||||||||||||||||||| ||||||||||||||||||||||||||
Db 1190 GCAGAGGTTGCAGTGAGCCGAGATCACGCCACTGCACTCCAGCCTGGGCAACAGAGCGAG 1131

Qy 741 ACTCCAGCTTAAAAAAAAAAAAAAAAAAAAAAAGGAGA 780
      ||||| ||||||||||||||| ||||| || ||| |||
Db 1130 ACTCCTTCTCAGAAAAAAAAAAATAATAATAAAATA 1091

RESULT 12
US-09-735-934A-3
; Sequence 3, Application US/09735934A
; Patent No. 6372468
; GENERAL INFORMATION:
; APPLICANT: LI, Jiayin et al
; TITLE OF INVENTION: ISOLATED HUMAN KINASE PROTEINS, NUCLEIC
; TITLE OF INVENTION: ACID MOLECULES ENCODING HUMAN KINASE PROTEINS, AND USES
; TITLE OF INVENTION: THEREOF
; FILE REFERENCE: CL000851
; CURRENT APPLICATION NUMBER: US/09/735,934A
; CURRENT FILING DATE: 2000-12-14
; NUMBER OF SEQ ID NOS: 4
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 3
; LENGTH: 43950
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-735-934A-3

Query Match   23.6%;   Score 236.4;   DB 4;   Length 43950;
Best Local Similarity   80.7%;   Pred. No. 1.5e-55;
Matches 276;   Conservative 0;   Mismatches 66;   Indels 0;   Gaps 0;

Qy 443 TACACCCCTCTCAAGAGACAGTCATTGGCTGGGCACGGTGGGCTCATGCCTGCAATCCCAG 502
      ||||| || || |||| |||| ||||||||||||||||||||||||||||||||||||||
Db 37638 TGCAGCCATTAACAATGAGGGTATGGGGCCCAGGTGGTGTCATGCTTGTAATCCAG 37697

Qy 503 CACCTTGGGAGGCTGAGGCAGGAGGATGAGGAGTTGAGGACCAGCCTGGC 562
      ||| ||||||||||||||||||||||||||||||||| |||||||||||
Db 37698 CACTTTGGGGAGGCTGAGGCAGGAGGAACACCTGAGGTCAGGAGTTGGGACCAGCCTGGC 37757

```
Qy    563 CAATATGGCGAAACCTCATTTCTACTAAAAATACAAAAAACTAACCAGCGTGTGCTTG 622
          |||  |||  ||||||||||||||||||||||||||||||  |||| |||| ||| |||
Db  37758 TGATGTCGTGAAATGTCGTCTCTTACTAAAAATACAAAAATTACCCAGCCATGTGGCCGGG 37817

Qy    623 TGCCTGTAATCCCAGCTACTCGGGAGGCTGAGGCAGGAGAATCGCTTGAACCGGGGAGGC 682
          | |||||  ||||||||| |||||||||||||||||||||||| ||||||||| ||||||
Db  37818 TCCCTGTGATCCCAGCTATTCGGGAGGCTGAGGCAGGAGAATTGCTTGAACCCGGGAGGC 37877

Qy    683 AGAGGTTGCAGTGTGAGCTGAGATCACACCACTGCACTCCAGCCTGGGCGACAGAGCGAGAC 742
          |||||||||||| | |||| |||||||| | |||||||||||||||||||||||||||||
Db  37878 AGAGGTTGCAGTGAGCCAAGATCACGCTACTGCACTCCAGCCTGGGCGACAGAGCGAGAC 37937

Qy    743 TCCAGCTTAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAGGAGGACCAT 784
          |||  || ||||||||      ||| |||  ||||  || |  || ||||
Db  37938 TCCGTCTCAAAAAAAATAAAACAATGAAGGAAAGGTAGGCAT 37979
```

RESULT 13
US-08-480-784-20/c
; Sequence 20, Application US/08480784
; Patent No. 5693473
; GENERAL INFORMATION:
;   APPLICANT: Skolnick, Mark H.
;   APPLICANT: Goldgar, David E.
;   APPLICANT: Miki, Yoshio
;   APPLICANT: Swenson, Jeff
;   APPLICANT: Kamb, Alexander
;   APPLICANT: Harshman, Keith D.
;   APPLICANT: Shattuck-Eidens, Donna M.
;   APPLICANT: Tavtigian, Sean V.
;   APPLICANT: Wiseman, Roger W.
;   APPLICANT: Futreal, P. Andrew
;   TITLE OF INVENTION: 17q-Linked Breast and Ovarian Cancer
;   Susceptibility Gene
;   NUMBER OF SEQUENCES: 85
;   CORRESPONDENCE ADDRESS:
;   ADDRESSEE: Venable, Baetjer, Howard & Civiletti, LLP
;   STREET: 1201 New York Avenue, N.W., Suite 1000
;   CITY: Washington
;   STATE: DC
;   COUNTRY: USA
;   ZIP: 20005
;   COMPUTER READABLE FORM:
;   MEDIUM TYPE: Floppy disk
;   COMPUTER: IBM PC compatible
;   OPERATING SYSTEM: PC-DOS/MS-DOS
;   SOFTWARE: PatentIn Release #1.0, Version #1.30
;   CURRENT APPLICATION DATA:
;   APPLICATION NUMBER: US/08/480,784
;   FILING DATE:
;   CLASSIFICATION: 435
;   PRIOR APPLICATION DATA:
;   APPLICATION NUMBER: US 08/409,305
;   FILING DATE: 24-MAR-1995
;   PRIOR APPLICATION DATA:
;   APPLICATION NUMBER: US 08/348,824
;   FILING DATE: 29-NOV-1994
;   PRIOR APPLICATION DATA:
;   APPLICATION NUMBER: US 08/308,104
;   FILING DATE: 16-SEP-1994
;   PRIOR APPLICATION DATA:
;   APPLICATION NUMBER: US 08/300,266
;   FILING DATE: 02-SEP-1994
;   PRIOR APPLICATION DATA:
;   APPLICATION NUMBER: US 08/289,221
;   FILING DATE: 12-AUG-1994
;   ATTORNEY/AGENT INFORMATION:
;   NAME: Ihnen, Jeffrey L.
;   REGISTRATION NUMBER: 28,957
;   REFERENCE/DOCKET NUMBER: 24884-109347
;   TELECOMMUNICATION INFORMATION:
;   TELEPHONE: 202-962-4810
;   TELEFAX: 202-962-8300

; INFORMATION FOR SEQ ID NO: 20:
;   SEQUENCE CHARACTERISTICS:
;   LENGTH: 6769 base pairs
;   TYPE: nucleic acid
;   STRANDEDNESS: double
;   TOPOLOGY: linear
;   MOLECULE TYPE: DNA (genomic)
;   HYPOTHETICAL: NO
;   ANTI-SENSE: NO
;   ORIGINAL SOURCE:
;   ORGANISM: Homo sapiens
US-08-480-784-20

Query Match   23.6%; Score 236.2; DB 1; Length 6769;
Best Local Similarity   84.7%; Pred. No. 7.2e-56;
Matches 265; Conservative   0; Mismatches   48; Indels   0; Gaps   0;

Qy   468 TGGCTGGGCACGGTGGCTCATGCCTGCAATCCCAGCACCCTTGGGAGGCTGAGGCAGGCAG 527
         | |||||||||||||||||||||||||||||||||||||||||||||||| |||| | |
Db 5745 TTGCTGGGCACGGTGGCTCATGCCTGCAATCCCAGCACCCTTGGGAGGCCGCGTGAGGGTGG 5686

Qy   528 ATGAGCTTGAGGCCAGGAGTTCGAGACCAGCCTGGCCAATATGGCGGAAACCTCATTTCTAC 587
         | ||| ||||| |||||||||| || |||||||||||||| | || ||||| || ||||||
Db 5685 ACCACTTGAGGTCATGAGTTCAAGACCAGCCTGGCCAACATGGTGAAAACCCCATCTCTAC 5626

Qy   588 TAAAAATACAAAAAACTAACCAGGCGTGGTGGCGCGTGCCTGTAGTCCCAGCTACTCGGGA 647
         |||||||||||||| ||||||||||||||||||| ||||||||||||||||||||| ||
Db 5625 TAAAAATACAAAAAATTAGCCGGGCGTGGTGGCGCGCGCCTGTAGTCCCAGCTACTAGGAGA 5566

Qy   648 GGCTGAGGCAGGAGAATCGCTTGAACCGGGAGGCAGGAGGTTGCAGTGAGCTGAGATCAC 707
         ||||||||||||||| | || |||| ||||||||| |||||| | |||||||||||
Db 5565 GGCTGAGGCAGGAGAGTATGGCTTCAACCTGGGAGGCAGCGTTGCGGAGGAGATCAC 5506

Qy   708 ACCACTGCACTCCAGCCTGGGCGACAGAGCGAGACTCCGTCTCAAAAAAAAAAAAAAA 767
         ||| ||| |||||||||||||| ||||||| |||||| ||||||||||||||||||||
Db 5505 ACCCTGTACTCCAGCCTGGGCAACAGAGCAAGAACTCTGTCTCAAAAAAAAAAAAAAA 5446

Qy   768 AAAAAAAAGGAGA 780
         | || |||| |||
Db 5445 AGAGAAAAGAAA 5433

RESULT 14
US-08-483-553-20/c
; Sequence 20, Application US/08483553
; Patent No. 5709999
; GENERAL INFORMATION:
;   APPLICANT: Skolnick, Mark H.
;   APPLICANT: Goldgar, David E.
;   APPLICANT: Miki, Yoshio
;   APPLICANT: Swenson, Jeff
;   APPLICANT: Kamb, Alexander
;   APPLICANT: Harshman, Keith D.
;   APPLICANT: Shattuck-Eidens, Donna M.
;   APPLICANT: Tavtigian, Sean V.
;   APPLICANT: Wiseman, Roger W.
;   APPLICANT: Futreal, P. Andrew
;   TITLE OF INVENTION: 17q-Linked Breast and Ovarian Cancer
;   Susceptibility Gene
;   NUMBER OF SEQUENCES: 85
;   CORRESPONDENCE ADDRESS:
;   ADDRESSEE: Venable, Baetjer, Howard & Civiletti, LLP
;   STREET: 1201 New York Avenue, N.W., Suite 1000
;   CITY: Washington
;   STATE: DC
;   COUNTRY: USA
;   ZIP: 20005
;   COMPUTER READABLE FORM:
;   MEDIUM TYPE: Floppy disk
;   COMPUTER: IBM PC compatible
;   OPERATING SYSTEM: PC-DOS/MS-DOS
;   SOFTWARE: PatentIn Release #1.0, Version #1.30
;   CURRENT APPLICATION DATA:
```

; APPLICATION NUMBER: US/08/483,553
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/409,305
; FILING DATE: 24-MAR-1995
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/348,824
; FILING DATE: 29-NOV-1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/348,824
; FILING DATE: 29-NOV-1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/308,104
; FILING DATE: 16-SEP-1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/300,266
; FILING DATE: 02-SEP-1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/289,221
; FILING DATE: 12-AUG-1994
; ATTORNEY/AGENT INFORMATION:
; NAME: Ihnen, Jeffrey L.
; REGISTRATION NUMBER: 28,957
; REFERENCE/DOCKET NUMBER: 24884-109347
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 202-962-4810
; TELEFAX: 202-962-8300
INFORMATION FOR SEQ ID NO: 20:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 6769 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; ORIGINAL SOURCE:
; ORGANISM: Homo sapiens
US-08-483-553-20

Query Match   23.6%;   Score 236.2;   DB 1;   Length 6769;
Best Local Similarity 84.7%;   Pred. No. 7.2e-56;
Matches 265;   Conservative   0;   Mismatches 48;   Indels   0;   Gaps   0;

Qy   468 TGGCTGGGCACGGTGGCTCATGCCTGCAATCCAGCACCTTGGGAGGCTGAGGCAGGCAG 527
          ||||||||||||||||||||||||||||||| ||| |||||||||||||||||||||| ||
Db   5745 TTGCTGGGCACGGTGGCTCAAGCCTGTAATTCCAGCACTTTGGGAGGCTGAGGCGGGTGG 5686

Qy   528 ATGACTTGAGGCCAGGAGTTCGAGACCAGCCTGGCCAACATGGCGAAACCTCATTTCTAC 587
          ||||||| || ||||||||||||||| ||||||||||||||| ||||||| ||||||||
Db   5685 ACCACTTGAGGTCATGAGTTCAAGACCAGCAGCCTGCCAACATGGTGAAACCCCATCTCTAC 5626

Qy   588 TAAAAATACAAAAAACTAACCAGGCGTGGTGGCTTGTGCCTGTAATCCCAGCTACTCGGGA 647
          |||||||||||||||| ||||||||||||||||| |||||||||||||||| ||||| ||||
Db   5625 TAAAAATACAAAAAATTAGCGGGTGTTGGTGGCGCGCGCCTGTAGTCCCAGCTACTCAGAGA 5566

Qy   648 GGCTGAGGCAGGAGAATCGCTTGAACCGGGGAGGCGGAGGTTCAGTGAGCTGAGATCAC 707
          ||||||||||||||||||||||||| ||||||||||||||||| |||| ||| |||||||||
Db   5565 GGCTGAGGCAGGAGGATGGCGTGAGCCCGGAGGTGGAGCTTGCAGTGAGCCGAGATCAC 5506

Qy   708 ACCACTGCACTCCAGCCTGGGCGACAGAGCGAGACTCCAGCTTAAAAAAAAAAAAAAAAA 767
          ||||||||||||||||||||| || ||||||||||||| ||||||||||||||||||||
Db   5505 ACCTCTGTACTCCAGCCTGGGCAACAGAGCAAGACTCTGTCTCAAAAAAAAAAAAAAAAA 5446

Qy   768 AAAAAAAAGGAGA 780
          || ||| || |||
Db   5445 AGAGAGAAAGAA 5433

; APPLICANT: Simard, Jacques
; APPLICANT: Emi, Mitsuru
; APPLICANT: Nakamura, Yusuke
; APPLICANT: Durocher, Francine
; TITLE OF INVENTION: 17q-Linked Breast and Ovarian Cancer
; TITLE OF INVENTION: Susceptibility Gene
; NUMBER OF SEQUENCES: 85
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Venable, Baetjer, Howard & Civiletti, LLP
; STREET: 1201 New York Avenue, N.W., Suite 1000
; CITY: Washington
; STATE: DC
; COUNTRY: USA
; ZIP: 20005
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/487,002
; FILING DATE:
; CLASSIFICATION: 424
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/409,305
; FILING DATE: 24-MAR-1995
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/348,824
; FILING DATE: 29-NOV-1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/348,824
; FILING DATE: 29-NOV-1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/308,104
; FILING DATE: 16-SEP-1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/300,266
; FILING DATE: 02-SEP-1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/289,221
; FILING DATE: 12-AUG-1994
; ATTORNEY/AGENT INFORMATION:
; NAME: Ihnen, Jeffrey L.
; REGISTRATION NUMBER: 28,957
; REFERENCE/DOCKET NUMBER: 24884-109347
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 202-962-4810
; TELEFAX: 202-962-8300
INFORMATION FOR SEQ ID NO: 20:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 6769 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; ORIGINAL SOURCE:
; ORGANISM: Homo sapiens
US-08-487-002-20

Query Match   23.6%;   Score 236.2;   DB 1;   Length 6769;
Best Local Similarity 84.7%;   Pred. No. 7.2e-56;
Matches 265;   Conservative   0;   Mismatches 48;   Indels   0;   Gaps   0;

Qy   468 TGGCTGGGCACGGTGGCTCATGCCTGCAATCCCAGCACCTTGGGAGGCTGAGGCAGGCAG 527
          ||||||||||||||||||||||||||||| ||| |||||||||||||||||||||| ||
Db   5745 TTGCTGGGCACGGTGGCTCAAGCCTGTAATTCCAGCACTTTGGGAGGCTGAGGCGGGTGG 5686

Qy   528 ATGACTTGAGGCCAGGAGTTCGAGACCAGCCTGCCAATATGGCGAAACCTCATTTCTAC 587
          ||||||| || ||||||||||||||| |||||||||||||| ||||||| ||||||||
Db   5685 ACCACTTGAGGTCATGAGTTCAAGACCAGCAGCCTGCCAACATGGTGAAACCCCATCTCTAC 5626

Qy   588 TAAAAATACAAAAAACTAACCAGGCGTGGTGGCTTGTGCCTGTAATCCCAGCTACTCGGGA 647
          |||||||||||||||| ||||||||||||| ||||||| |||||||||| ||||| ||||
Db   5625 TAAAAATACAAAAAATTAGCGGGTGTTGGTGGCGCGCGCCTGTAGTCCCAGCTACTCAGAGA 5566

```
Qy    648 GGCTGAGGCAGGAGGAGAATCGCTTGAACCGGGGAGGCAGAGGTTGCAGTGAGCTGAGATCAC 707
          |||||||||||||| || |||| |||| |||||||||| | ||| ||||||||
Db   5565 GGCTGAGGCAGGAGTATGGCTTCAACCTGGGAGGCAGACGTTGCGGAGGTGAGATCAC 5506

Qy    708 ACCACTGCACTCCAGCCTGGGCGACAGAGCGAGACTCCAGCTTTAAAAAAAAAAAAAA 767
          ||| ||| ||||||||||||| ||||||| |||||||||||||||||||||
Db   5505 ACCTCTGTACTCCAGCCTGGGCAACAGAGCAAGACTCTGTCTCAAAAAAAAAAAAAAA 5446

Qy    768 AAAAAAAAGGGAGA 780
          | | | ||| || |
Db   5445 AGAGAGAAAGAAA 5433
```

Search completed: March 29, 2003, 22:56:01
Job time : 953.481 secs

OM nucleic - nucleic search, using sw model
GenCore version 5.1.4_p5_4578
Copyright (c) 1993 - 2003 Compugen Ltd.

Run on: March 29, 2003, 21:53:16 ; Search time 105.412 seconds
                 (without alignments)
                 8080.893 Million cell updates/sec

Title: 1001
Perfect score: 1001
Sequence: 1 ttactttaccatcctgggg..........ggcccatgggtgaggacatg 1001

Scoring table: IDENTITY_NUC
                 Gapop 10.0 , Gapext 1.0

Searched: 574371 seqs, 425486471 residues

Total number of hits satisfying chosen parameters: 1148742

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
                 Maximum Match 100%
                 Listing first 45 summaries

Database :   Published_Applications_NA:*
                 1: /cgn2_6/ptodata/2/pubpna/US07_PUBCOMB.seq:*
                 2: /cgn2_6/ptodata/2/pubpna/PCT_NEW_PUB.seq:*
                 3: /cgn2_6/ptodata/2/pubpna/US06_NEW_PUB.seq:*
                 4: /cgn2_6/ptodata/2/pubpna/US06_PUBCOMB.seq:*
                 5: /cgn2_6/ptodata/2/pubpna/US07_NEW_PUB.seq:*
                 6: /cgn2_6/ptodata/2/pubpna/PCTUS_PUBCOMB.seq:*
                 7: /cgn2_6/ptodata/2/pubpna/US08_NEW_PUB.seq:*
                 8: /cgn2_6/ptodata/2/pubpna/US08_PUBCOMB.seq:*
                 9: /cgn2_6/ptodata/2/pubpna/US09_NEW_PUB.seq:*
                10: /cgn2_6/ptodata/2/pubpna/US09_PUBCOMB.seq:*
                11: /cgn2_6/ptodata/2/pubpna/US10_NEW_PUB.seq:*
                12: /cgn2_6/ptodata/2/pubpna/US10_PUBCOMB.seq:*
                13: /cgn2_6/ptodata/2/pubpna/US60_NEW_PUB.seq:*
                14: /cgn2_6/ptodata/2/pubpna/US60_PUBCOMB.seq:*

SUMMARIES

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

Result No.	Score	% Query Match	Match Length	DB	ID	Description
1	250.2	25.0	40433	10	US-09-880-107-3327	Sequence 3327, Ap
2	248	24.8	4374	9	US-10-028-072-125	Sequence 125, App
3	248	24.8	4374	9	US-10-121-049-125	Sequence 125, App
4	248	24.8	4374	9	US-10-123-904-125	Sequence 125, App
5	248	24.8	4374	9	US-10-140-470-125	Sequence 125, App
6	248	24.8	4374	9	US-10-175-746-125	Sequence 125, App
7	248	24.8	4374	9	US-10-176-918-125	Sequence 125, App
8	248	24.8	4374	9	US-10-176-921-125	Sequence 125, App
9	248	24.8	4374	9	US-10-137-865-125	Sequence 125, App
10	248	24.8	4374	9	US-10-140-474-125	Sequence 125, App
11	248	24.8	4374	9	US-10-142-431-125	Sequence 125, App
12	248	24.8	4374	9	US-10-143-114-125	Sequence 125, App
13	248	24.8	4374	9	US-10-140-002-125	Sequence 125, App
14	248	24.8	4374	9	US-10-142-419-125	Sequence 125, App
15	248	24.8	4374	9	US-10-123-262-125	Sequence 125, App
16	248	24.8	4374	9	US-10-142-423-125	Sequence 125, App
17	248	24.8	4374	9	US-10-121-050-125	Sequence 125, App
18	248	24.8	4374	9	US-10-141-755-125	Sequence 125, App
19	247.8	24.8	4766	9	US-09-764-868-1443	Sequence 1443, Ap
20	247.8	24.8	65464	9	US-09-859-888-3	Sequence 3, Appli
21	245.2	24.5	6892	10	US-09-764-877-3770	Sequence 3770, Ap
22	244.2	24.4	1912	10	US-09-800-729-32	Sequence 32, Appl
23	244.2	24.4	1332	12	US-10-080-644-1	Sequence 1, Appli
24	244	24.2	14012	9	US-09-819-994-3	Sequence 3, Appli
25	244	24.3	170834	10	US-09-835-232-7	Sequence 7, Appli
26	244	24.3	174424	10	US-09-982-768A-314	Sequence 314, App
27	243.2	24.3	58837	10	US-09-967-768A-116	Sequence 116, Appl
28	243	24.3	199997	10	US-09-822-246-3	Sequence 3220, Ap
29	242.2	24.2	41936	10	US-09-764-877-3220	Sequence 3220, Ap
30	242.2	24.2	19616	10	US-09-854-883-243	Sequence 243, App
31	242	24.2	75899	10	US-09-820-002-3	Sequence 3, Appli
32	241.6	24.1	21784	10	US-09-967-768A-2419	Sequence 2419, Ap
33	241.6	24.1	24699	10	US-09-764-860-797	Sequence 797, App
34	241.4	24.1	32146	10	US-09-764-877-2680	Sequence 2680, Ap
35	241.2	24.1	12932	10	US-10-092-154-1132	Sequence 1132, Ap
36	241.2	24.1	12932	9	US-09-764-847-1132	Sequence 1132, Ap
37	241	24.1	20247	10	US-09-764-877-1132	Sequence 314, Ap
38	240.8	24.1	174424	9	US-10-092-154-314	Sequence 314, Ap
39	240.8	24.1	4622	10	US-09-764-877-1141	Sequence 1141, Ap
40	240.8	24.0	4622	10	US-09-764-847-1141	Sequence 1141, Ap
41	240.6	24.0	6074	10	US-09-880-107-2378	Sequence 2378, Ap
42	240.6	24.0	8095	10	US-09-880-107-2228	Sequence 2228, Ap
43	240.6	24.0	35641	10	US-09-962-436-306	Sequence 306, App
44	240.2	24.0	35641	10	US-09-880-107-2225	Sequence 2225, Ap
45	240.2	24.0	110096	10	US-09-880-107-1542	Sequence 1542, Ap

ALIGNMENTS

RESULT 1
US-09-880-107-3327/c

Sequence 3327, Application US/09880107
; Patent No. US2002014298IA1
; GENERAL INFORMATION:
US-09-880-107-3327
; APPLICANT: Horne, Darci T.
; APPLICANT: Vockley, Joseph G.
; APPLICANT: Scherf, Uwe
; APPLICANT: Gene Logic, Inc.
; TITLE OF INVENTION: Gene Expression Profiles in Liver Cancer
; FILE REFERENCE: 44921-5028-WO
; CURRENT APPLICATION NUMBER: US/09/880,107
; CURRENT FILING DATE: 2001-06-14
; PRIOR APPLICATION NUMBER: US 60/211,379
; PRIOR FILING DATE: 2000-06-14
; PRIOR APPLICATION NUMBER: 2000-06-14
; PRIOR APPLICATION NUMBER: US 60/237,054
; PRIOR FILING DATE: 2000-10-02
; NUMBER OF SEQ ID NOS: 3950
; SOFTWARE: Patentin Ver. 2.1
; SEQ ID NO 3327
; LENGTH: 40433
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: Genbank Accession No. US2002014298IA1 U22376
US-09-880-107-3327

Query Match 25.0%;   Score 250.2;   DB 10;   Length 40433;
Best Local Similarity 77.5%;
Matches 303;   Conservative 0;   Pred. No. 3.7e-64;
Matches 303;   Conservative 0;   Mismatches 88;   Indels 0;   Gaps 0;

```
Qy  466 ATTGGCTGGGCACGGTGGCTCATGCCTCATTGCAAATCCCAGCACCTTTGGGAGGCTGAGGCAGGC  525
        |||||| |||||| |||||||||||||||| |||||||| |||||||||| ||||||||||| |||
Db 9505 ACTGGCTGCGGGCGACGGTGGCTCATGCCTGTGAATCCTAGCACTTCGGAGGCCGAGGTGGGC  9446

Qy  526 AGATGACTTGAGGCCAGGAGTTCGAGACCAGCCTGGCCAACATGTGGCCAAACCTCATTTCT  585
        |||| ||||||||| |||||||| | |||||||||||| ||||||| ||| ||||| | ||||
Db 9445 AGATCACTTGAGGTCAGGAGTTTGGGACCAGCCTGGGCCAACATGGTGAAACCCTTGTCTCT  9386

Qy  586 ACTAAAAATACAAAAAAAACTACCAGGCGTGGTGGTTGTGCCTGTAATCCCAGCTACTCGG  645
        ||||||||||||||||||  ||  |||||||||||| | | |||||| |||||| ||||| |
Db 9385 ACTAAAAATACAAAAAATTAGCCAGGCATGGTGGCGGGCGCACCTGTGATCCCATCTACTAAG  9326
```

```
Qy    646  GAGGCTGAGGCAGGAGAATCGTTGAACCGGGGAGGCGGAGAGGTTGCAGTGAGCTGAGATC    705
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   9325  GGGGCTGAGGCACGAGAATCGCTTGAACCTGGGAGGCGGAGAGTTTGCAGTGAGCTGAGATC   9266

Qy    706  ACACCACTGCACTCCAGCCTGGGCGACGAGCGCAGAGCGACTCCAGCTTAAAAAAAAAAA    765
           ||||||||||||||||||||||||||||||||||||||||||||||||
Db   9265  GTGCCACTGCACTCCAGCCTGGGCAACAAAGCTAGAATTCCATTCTCAAAAAAAAAAAAACA   9206

Qy    766  AAAAAAAAAAGGAGACCATCACTGCTGTCCTGCCATTCTTACAGATGAAAAAACAGGCTCA    825
           ||||||||  |||  |||  |  ||| ||
Db   9205  AAAAACAAAAGAAAAAAAGGAAAAACTACAAGTGGGAAGCCAGTTAG   9146

Qy    826  GAGGTTGAATCGTTTTCCTGAAGTCAGACAG    856
           ||||  |||   || | |||  | |  |||
Db   9145  GAAGTTATATGATAGTCCAAATGAGAGGCAG   9115
```

RESULT 2
US-10-028-072-125
Sequence 125, Application US/10028072
Publication No. US20030004311A1
GENERAL INFORMATION:

```
;  APPLICANT:  Baker,Kevin P.
;  APPLICANT:  Beresini,Maureen
;  APPLICANT:  DeForge,Laura
;  APPLICANT:  Desnoyers,Luc
;  APPLICANT:  Filvaroff,Ellen
;  APPLICANT:  Gao,Wei-Qiang
;  APPLICANT:  Gerritsen,Mary E.
;  APPLICANT:  Goddard,Audrey
;  APPLICANT:  Godowski,Paul J.
;  APPLICANT:  Gurney,Austin L.
;  APPLICANT:  Sherwood,Steven
;  APPLICANT:  Smith,Victoria
;  APPLICANT:  Stewart,Timothy A.
;  APPLICANT:  Tumas,Daniel
;  APPLICANT:  Watanabe,Colin K
;  APPLICANT:  Wood,William
;  APPLICANT:  Zhang
;  TITLE OF INVENTION:
;  FILE REFERENCE:
;  CURRENT APPLICATION NUMBER: US/10,028,072
;  CURRENT FILING DATE:  2001-12-19
;  PRIOR APPLICATION NUMBER: 60/049911
;  PRIOR FILING DATE:  1997-06-18
;  PRIOR APPLICATION NUMBER: 60/056974
;  PRIOR FILING DATE:  1997-08-26
;  PRIOR APPLICATION NUMBER: 60/059113
;  PRIOR FILING DATE:  1997-09-17
;  PRIOR APPLICATION NUMBER: 60/059115
;  PRIOR FILING DATE:  1997-09-17
;  PRIOR APPLICATION NUMBER: 60/059117
;  PRIOR FILING DATE:  1997-09-17
;  PRIOR APPLICATION NUMBER: 60/059122
;  PRIOR FILING DATE:  1997-09-17
;  PRIOR APPLICATION NUMBER: 60/059184
;  PRIOR FILING DATE:  1997-09-17
;  PRIOR APPLICATION NUMBER: 60/059263
;  PRIOR FILING DATE:  1997-09-18
;  PRIOR APPLICATION NUMBER: 60/059352
;  PRIOR FILING DATE:  1997-09-19
;  PRIOR APPLICATION NUMBER: 60/059588
;  PRIOR FILING DATE:  1997-09-19
;  PRIOR APPLICATION NUMBER: 60/059836
;  PRIOR FILING DATE:  1997-09-24
;  PRIOR APPLICATION NUMBER: 60/062250
;  PRIOR FILING DATE:  1997-10-17
;  PRIOR APPLICATION NUMBER: 60/062285
;  PRIOR FILING DATE:  1997-10-17
;  PRIOR APPLICATION NUMBER: 60/062287
;  PRIOR FILING DATE:  1997-10-17
;  PRIOR APPLICATION NUMBER: 60/062814
```

```
;  PRIOR FILING DATE:  1997-10-24
;  PRIOR APPLICATION NUMBER: 60/062816
;  PRIOR FILING DATE:  1997-10-24
;  PRIOR APPLICATION NUMBER: 60/063045
;  PRIOR FILING DATE:  1997-10-24
;  PRIOR APPLICATION NUMBER: 60/063082
;  PRIOR FILING DATE:  1997-10-31
;  PRIOR APPLICATION NUMBER: 60/063127
;  PRIOR FILING DATE:  1997-10-24
;  PRIOR APPLICATION NUMBER: 60/063327
;  PRIOR FILING DATE:  1997-10-27
;  PRIOR APPLICATION NUMBER: 60/063329
;  PRIOR FILING DATE:  1997-10-27
;  PRIOR APPLICATION NUMBER: 60/063550
;  PRIOR FILING DATE:  1997-10-28
;  PRIOR APPLICATION NUMBER: 60/063561
;  PRIOR FILING DATE:  1997-10-27
;  PRIOR APPLICATION NUMBER: 60/063704
;  PRIOR FILING DATE:  1997-10-29
;  PRIOR APPLICATION NUMBER: 60/063733
;  PRIOR FILING DATE:  1997-10-29
;  PRIOR APPLICATION NUMBER: 60/063735
;  PRIOR FILING DATE:  1997-10-29
;  PRIOR APPLICATION NUMBER: 60/063738
;  PRIOR FILING DATE:  1997-10-29
;  PRIOR APPLICATION NUMBER: 60/063755
;  PRIOR FILING DATE:  1997-10-17
;  PRIOR APPLICATION NUMBER: 60/064248
;  PRIOR FILING DATE:  1997-11-03
;  PRIOR APPLICATION NUMBER: 60/064809
;  PRIOR FILING DATE:  1997-11-07
;  PRIOR APPLICATION NUMBER: 60/055186
;  PRIOR FILING DATE:  1997-11-12
;  PRIOR APPLICATION NUMBER: 60/065846
;  PRIOR FILING DATE:  1997-11-17
;  PRIOR APPLICATION NUMBER: 60/066364
;  PRIOR FILING DATE:  1997-11-21
;  PRIOR APPLICATION NUMBER: 60/066453
;  PRIOR FILING DATE:  1997-11-24
;  PRIOR APPLICATION NUMBER: 60/066511
;  PRIOR FILING DATE:  1997-11-24
;  PRIOR APPLICATION NUMBER: 60/066770
;  PRIOR FILING DATE:  1997-11-24
;  PRIOR APPLICATION NUMBER: 60/069212
;  PRIOR FILING DATE:  1997-12-11
;  PRIOR APPLICATION NUMBER: 60/069278
;  PRIOR FILING DATE:  1997-12-11
;  PRIOR APPLICATION NUMBER: 60/069334
;  PRIOR FILING DATE:  1997-12-11
;  PRIOR APPLICATION NUMBER: 60/069694
;  PRIOR FILING DATE:  1997-12-16
;  PRIOR APPLICATION NUMBER: 60/072320
;  PRIOR FILING DATE:  1998-01-23
;  PRIOR APPLICATION NUMBER: 60/073612
;  PRIOR FILING DATE:  1998-02-04
;  PRIOR APPLICATION NUMBER: 60/074086
;  PRIOR FILING DATE:  1998-02-09
;  PRIOR APPLICATION NUMBER: 60/074092
;  PRIOR FILING DATE:  1998-02-09
;  PRIOR APPLICATION NUMBER: 60/077791
;  PRIOR FILING DATE:  1998-03-12
;  PRIOR APPLICATION NUMBER: 60/078910
;  PRIOR FILING DATE:  1998-03-20
;  PRIOR APPLICATION NUMBER: 60/079294
;  PRIOR FILING DATE:  1998-03-25
;  PRIOR APPLICATION NUMBER: 60/079663
;  PRIOR FILING DATE:  1998-02-27
;  PRIOR APPLICATION NUMBER: 60/079728
;  PRIOR FILING DATE:  1998-03-27
;  PRIOR APPLICATION NUMBER: 60/080165
;  PRIOR FILING DATE:  1998-03-31
;  PRIOR APPLICATION NUMBER: 60/081203
;  PRIOR FILING DATE:  1998-04-09
```

PRIOR APPLICATION NUMBER: 60/081229
PRIOR FILING DATE: 1998-04-09
PRIOR APPLICATION NUMBER: 60/081695
PRIOR FILING DATE: 1998-04-14
PRIOR APPLICATION NUMBER: 60/081817
PRIOR FILING DATE: 1998-04-15
PRIOR APPLICATION NUMBER: 60/081818
PRIOR FILING DATE: 1998-04-15
PRIOR APPLICATION NUMBER: 60/082999
PRIOR FILING DATE: 1998-04-24
PRIOR APPLICATION NUMBER: 60/083322
PRIOR FILING DATE: 1998-04-28
PRIOR APPLICATION NUMBER: 60/083545
PRIOR FILING DATE: 1998-04-29
PRIOR APPLICATION NUMBER: 60/084600
PRIOR FILING DATE: 1998-05-07
PRIOR APPLICATION NUMBER: 60/084627
PRIOR FILING DATE: 1998-05-07
PRIOR APPLICATION NUMBER: 60/084637
PRIOR FILING DATE: 1998-05-07
PRIOR APPLICATION NUMBER: 60/085149
PRIOR FILING DATE: 1998-05-12
PRIOR APPLICATION NUMBER: 60/085323
PRIOR FILING DATE: 1998-05-13
PRIOR APPLICATION NUMBER: 60/085338
PRIOR FILING DATE: 1998-05-13
PRIOR APPLICATION NUMBER: 60/085339
PRIOR FILING DATE: 1998-05-13
PRIOR APPLICATION NUMBER: 60/085579
PRIOR FILING DATE: 1998-05-15
PRIOR APPLICATION NUMBER: 60/085697
PRIOR FILING DATE: 1998-05-15
PRIOR APPLICATION NUMBER: 60/085704
PRIOR FILING DATE: 1998-05-15
PRIOR APPLICATION NUMBER: 60/086414
PRIOR FILING DATE: 1998-05-22
PRIOR APPLICATION NUMBER: 60/086430
PRIOR FILING DATE: 1998-05-22
PRIOR APPLICATION NUMBER: 60/087106
PRIOR FILING DATE: 1998-05-28
PRIOR APPLICATION NUMBER: 60/088026
PRIOR FILING DATE: 1998-06-04
PRIOR APPLICATION NUMBER: 60/088730
PRIOR FILING DATE: 1998-06-10
PRIOR APPLICATION NUMBER: 60/088741
PRIOR FILING DATE: 1998-06-10
PRIOR APPLICATION NUMBER: 60/088810
PRIOR FILING DATE: 1998-06-10
PRIOR APPLICATION NUMBER: 60/088858
PRIOR FILING DATE: 1998-06-10
PRIOR APPLICATION NUMBER: 60/089532
PRIOR FILING DATE: 19/98-06-11
PRIOR APPLICATION NUMBER: 60/089599
PRIOR FILING DATE: 1998-06-17
PRIOR APPLICATION NUMBER: 60/089907
PRIOR FILING DATE: 1998-06-17
PRIOR APPLICATION NUMBER: 60/089947
PRIOR FILING DATE: 1998-06-18
PRIOR APPLICATION NUMBER: 60/090349
PRIOR FILING DATE: 1998-06-19
PRIOR APPLICATION NUMBER: 60/090429
PRIOR FILING DATE: 1998-06-23
PRIOR APPLICATION NUMBER: 60/090445
PRIOR FILING DATE: 1998-06-24
PRIOR APPLICATION NUMBER: 60/090538
PRIOR FILING DATE: 1998-06-24
PRIOR APPLICATION NUMBER: 60/090863
PRIOR FILING DATE: 1998-06-24
PRIOR APPLICATION NUMBER: 60/091360
PRIOR FILING DATE: 1998-06-26
PRIOR APPLICATION NUMBER: 60/091519
PRIOR FILING DATE: 1998-07-01
PRIOR APPLICATION NUMBER: 60/091982
PRIOR FILING DATE: 1998-07-02

PRIOR FILING DATE: 1998-07-07

RESULT 3
US-10-121-049-125
Sequence 125, Application US/10121049
Publication No. US20030222239A1
GENERAL INFORMATION:
    APPLICANT: Baker, Kevin P.
    APPLICANT: Beresini, Maureen
    APPLICANT: DeForge, Laura
    APPLICANT: Desnoyers, Luc
    APPLICANT: Filvaroff, Ellen
    APPLICANT: Gao, Wei-Qiang
    APPLICANT: Gerritsen, Mary E.
    APPLICANT: Goddard, Audrey
    APPLICANT: Godowski, Paul J.
    APPLICANT: Gurney, Austin L.
    APPLICANT: Sherwood, Steven
    APPLICANT: Smith, Victoria
    APPLICANT: Stewart, Timothy A.
    APPLICANT: Tumas, Daniel
    APPLICANT: Watanabe, Colin K
    APPLICANT: Wood, William
    APPLICANT: Zhang, Zemin
    TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
    TITLE OF INVENTION: ACIDS ENCODING THE SAME
    FILE REFERENCE: P330R1C17
    CURRENT APPLICATION NUMBER: US/10/121,049
    CURRENT FILING DATE: 2002-04-12
        Prior Application removed - See File Wrapper or Palm
    NUMBER OF SEQ ID NOS: 550
    SEQ ID NO 125
        LENGTH: 4374
        TYPE: DNA
        ORGANISM: Homo Sapien
US-10-121-049-125

Query Match   24.8%; Score 248; DB 9; Length 4374;
Best Local Similarity   76.2%; Pred. No. 6.2e-64;
Matches 305; Conservative 0; Mismatches 95; Indels 0; Gaps 0;

Qy   451 TCTCAAGAGACAGTCATTGGCTGGGGCGACGGTGGCTCATGCCTGCAATCCCAGCACCTTGG 510
          |||||||
Db   3112 TTTTAAAATATGTACTCTGGCCGGTTGCAGTGGCTCACGCCTGTAATCCCAGCACTTTGG 3171

Qy   511 GAGGCTGAGGCAGGCAGATGACTTGAGGCCAGGAGTTCGAGACCAGCCTGGGCAAATATGG 570
          |||| ||| ||| |||| ||||||| |||||||||||||||||||||| ||||
Db   3172 CAGGCCGAGATGGGCGGGATCTTTTGAGGTCAGGAGTTTGAGACCAGCCTGGGCAACATGG 3231

Qy   571 CGAAAACCTCATTTCTACTAAAAATACAAAAAACTAACCAGGCGTGGTGGTGGGCGCCTGTA 630
          ||||||| | ||| ||||||||||||||||||||||| |||| |||| ||||||| ||
Db   3232 TGAAAACCCCGTCTCTACTAAAAATACAAAAATTAGCCAGGCATGGTGGCGCATTTGCCTGTA 3291

Qy   631 ATCCCAGCTACTCCGGCAGGCGGCTGAGGCAGGAGAATCGCTTGAACCGGGGAGGCGGAGGTTG 690
          ||||||||||||| |||| ||||||||||| ||||||| |||||||||||| |||||||||||
Db   3292 ATCCCAGCTACTCGGGGAGGCTGAGGCAGGAGAATCACTTGAACCTGGGAGGCGGAGGTTG 3351

Qy   691 CAGTGAGCTGAGATCACACCACCACTGCACTCCAGCCTGGGCGACGAGAGCGAGACTCCAGCTT 750
          ||||||||||||||||| |||| ||||| ||| |||||||||||| |||||||| |||||||| ||
Db   3352 CAGTGAGCTGAGATTACACCACTGCACTCCAGCCTGGGTGACAGAGTGAGACTCCGTCTC 3411

Qy   751 AAAAAAAAAAAAAAAAAAAAAAAAAAGGAGACCAATCACTGCTTGTCCTGCATTTCTTACAGAT 810
          |||| ||||||||||||||||||||||| |||| |||||||| || ||||
Db   3412 AAAAACTGAAAAATAAAAATAAAAATATGTATTCTCCTAACTGAAAATATTTACTTAATCTG 3471

Qy   811 GAAAAAACAGGCTCAGAGAGGTTGAATCGTTTTCCTGAAGTC 850
          ||||| |||||||| || ||| ||| |||| ||||
Db   3472 GAAAACAATGTAACTATTTTTTAAAGTGGTTACATCTATTC 3511

RESULT 4
US-10-123-904-125
Sequence 125, Application US/10123904
Publication No. US20030022328A1
GENERAL INFORMATION:

; APPLICANT: Baker,Kevin P.
; APPLICANT: Beresini,Maureen
; APPLICANT: DeForge,Laura
; APPLICANT: Desnoyers,Luc
; APPLICANT: Filvaroff,Ellen
; APPLICANT: Gao,Wei-Qiang
; APPLICANT: Gerritsen,Mary E.
; APPLICANT: Goddard,Audrey
; APPLICANT: Godowski,Paul J.
; APPLICANT: Gurney,Austin L.
; APPLICANT: Sherwood,Steven
; APPLICANT: Smith,Victoria
; APPLICANT: Stewart,Timothy A.
; APPLICANT: Tumas,Daniel
; APPLICANT: Watanabe,Colin K
; APPLICANT: Wood,William
; APPLICANT: Zhang,Zemin
; TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
; TITLE OF INVENTION: ACIDS ENCODING THE SAME
; FILE REFERENCE: P3330R1C54
; CURRENT APPLICATION NUMBER: US/10/123,904
; CURRENT FILING DATE: 2002-04-16
; PRIOR APPLICATION NUMBER: US/10/123,904
; Prior Application removed - See File Wrapper or Palm
; NUMBER OF SEQ ID NOS: 550
; SEQ ID NO 125
; LENGTH: 4374
; TYPE: DNA
; ORGANISM: Homo Sapien
US-10-123-904-125

Query Match   24.8%; Score 248; DB 9; Length 4374;
Best Local Similarity   76.2%; Pred. No. 6.2e-64;
Matches   305; Conservative   0; Mismatches   95; Indels   0; Gaps   0;

Qy   451 TCTCAAGAGACAGTCATTGGCTGGGGCACGGTGGCTCATGCCTGCAATCCCAGCACCTTGG 510
          | |||| | | | ||||| || ||||||||||||||||||| || |||||||||||||||
Db   3112 TTTTAAATATGTACTCTCGGCCGGTTGCAGTGGCTCACGCCTGTAATCCCAGCACTTTGG 3171

Qy   511 GAGGCTGAGGCGAGGCGAGATGACTTGAGGCCAGGAGTTCGAGACCAGCCTGGCCAATATGG 570
          ||| ||||||| || || || ||| |||||||||||||||| || ||||||||||||||| |
Db   3172 CAGGCCGAGATGGGCGGGATCTTTTGAGGTCAGGAGTTTGAGACCAGCCATGGTGGCTGTA 3231

Qy   571 CGAAACCTCATTTCTACTAAAAATACAAAAAATTAACCAGGCGTGGTGGTGGGCTGTGTA 630
          |||||| | ||| ||||||||||||||||| |||||| |||| | | ||||| || ||
Db   3232 TGAAACCCCGTCTCTACTAAAAATACAAAAATTAGCCAGGCGATGGTGGGCCATTTGCCTGTA 3291

Qy   631 ATCCCAGCTACTCGGGAGGCTGAGGCAGGAGAATCGCTTGAACCGGGAGGCGGAGGTTG 690
          ||||||||| | ||| || ||||||||| ||||| |||| || ||| ||| |||||||||
Db   3292 ATCCCAGCTACTCGGGAGGCTGAGGCAGGAGAATCACTTGAACCTGGGAGGCGGAGGTTG 3351

Qy   691 CAGTGAGCTGAGATCACACACTGCACTCCAGCCTGGGCGACAGAGCGAGACTCCGTCTT 750
          |||||||||||||| ||| |||||||||||||||||| | ||||| ||| || || |||
Db   3352 CAGTGAGCTGAGATTACACCACTGCACTCCAGCCTGGGCGACAGAGTGAGACTCC 3411

Qy   751 AAAAAAAAACAGGCTCAGAGGTTGAATCGTTTTCCTGAAGTC 810
          ||||||||| ||||| ||| || || | ||| | ||||| |
Db   3412 AAAAACTGAAAATAAAAATATGTATTCTCCTAACTGAAAATATTTACTTAATCTG 3471

Qy   811 GAAAAACAGGCTCAGAGGTTGAATCCGTTTTTCCTGAAGTC 850
          ||||| | ||| ||| | | || | ||| |||||||| |
Db   3472 GAAAACAAGTAACTATTTTTAAAGTGGTTACATCTATTC 3511

RESULT 5
US-10-140-470-125
Sequence 125, Application US/10140470
Publication No. US20030022331A1
GENERAL INFORMATION:

; APPLICANT: Baker,Kevin P.
; APPLICANT: Beresini,Maureen
; APPLICANT: DeForge,Laura
; APPLICANT: Desnoyers,Luc
; APPLICANT: Filvaroff,Ellen
; APPLICANT: Gao,Wei-Qiang
; APPLICANT: Gerritsen,Mary E.
; APPLICANT: Goddard,Audrey
; APPLICANT: Godowski,Paul J.
; APPLICANT: Gurney,Austin L.
; APPLICANT: Sherwood,Steven
; APPLICANT: Smith,Victoria
; APPLICANT: Stewart,Timothy A.
; APPLICANT: Tumas,Daniel
; APPLICANT: Watanabe,Colin K
; APPLICANT: Wood,William
; APPLICANT: Zhang,Zemin
; TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
; TITLE OF INVENTION: ACIDS ENCODING THE SAME
; FILE REFERENCE: P3330R1C160
; CURRENT APPLICATION NUMBER: US/10/140,470
; CURRENT FILING DATE: 2002-05-06
; Prior Application removed - See Palm or File Wrapper
; NUMBER OF SEQ ID NOS: 550
; SEQ ID NO 125
; LENGTH: 4374
; TYPE: DNA
; ORGANISM: Homo Sapien
US-10-140-470-125

Query Match   24.8%; Score 248; DB 9; Length 4374;
Best Local Similarity   76.2%; Pred. No. 6.2e-64;
Matches   305; Conservative   0; Mismatches   95; Indels   0; Gaps   0;

Qy   451 TCTCAAGAGACAGTCATTGGCTGGGGCACGGTGGCTCATGCCTGCAATCCCAGCACCTTGG 510
          | |||| | | | ||||| || ||||||||||||||||||| || |||||||||||||||
Db   3112 TTTTAAATATGTACTCTCGGCCGGTTGCAGTGGCTCACGCCTGTAATCCCAGCACTTTGG 3171

Qy   511 GAGGCTGAGGCGAGGCGAGATGACTTGAGGCCAGGAGTTCGAGACCAGCCTGGCCAATATGG 570
          ||| ||||||| || || || ||| |||||||||||||||| || ||||||||||||||| |
Db   3172 CAGGCCGAGATGGGCGGGATCTTTTGAGGTCAGGAGTTTGAGACCAGCCATGGTGGCTGTA 3231

Qy   571 CGAAACCTCATTTCTACTAAAAATACAAAAAATTAACCAGGCGTGGTGGTGGGCTGTGTA 630
          |||||| | ||| ||||||||||||||||| |||||| |||| | | ||||| || ||
Db   3232 TGAAACCCCGTCTCTACTAAAAATACAAAAATTAGCCAGGCGATGGTGGGCCATTTGCCTGTA 3291

Qy   631 ATCCCAGCTACTCGGGAGGCTGAGGCAGGAGAATCGCTTGAACCGGGAGGCGGAGGTTG 690
          ||||||||| | ||| || ||||||||| ||||| |||| || ||| ||| |||||||||
Db   3292 ATCCCAGCTACTCGGGAGGCTGAGGCAGGAGAATCACTTGAACCTGGGAGGCGGAGGTTG 3351

Qy   691 CAGTGAGCTGAGATCACACACTGCACTCCAGCCTGGGCGACAGAGCGAGACTCCGTCTT 750
          |||||||||||||| ||| |||||||||||||||||| | ||||| ||| || || |||
Db   3352 CAGTGAGCTGAGATTACACCACTGCACTCCAGCCTGGGCGACAGAGTGAGACTCC 3411

Qy   751 AAAAAAAAACAGGCTCAGAGGTTGAATCGTTTTCCTGAAGTC 810
          ||||||||| ||||| ||| || || | ||| | ||||| |
Db   3412 AAAAACTGAAAATAAAAATATGTATTCTCCTAACTGAAAATATTTACTTAATCTG 3471

Qy   811 GAAAAACAGGCTCAGAGGTTGAATCCGTTTTTCCTGAAGTC 850
          ||||| | ||| ||| | | || | ||| |||||||| |
Db   3472 GAAAACAAGTAACTATTTTTAAAGTGGTTACATCTATTC 3511

Query Match   24.8%; Score 248; DB 9; Length 4374;
Best Local Similarity   76.2%; Pred. No. 6.2e-64;
Matches   305; Conservative   0; Mismatches   95; Indels   0; Gaps   0;

Qy   451 TCTCAAGAGACAGTCATTGGCTGGGGCACGGTGGCTCATGCCTGCAATCCCAGCACCTTGG 510
          | |||| | | | ||||| || ||||||||||||||||||| || |||||||||||||||
Db   3112 TTTTAAATATGTACTCTCGGCCGGTTGCAGTGGCTCACGCCTGTAATCCCAGCACTTTGG 3171

Qy   571 CGAAACCTCATTTCTACTAAAAATACAAAAAATTAACCAGGCGTGGTGGTGGGCTGTA 630

RESULT 6

US-10-175-746-125
Sequence 125, Application US/10175746
Publication No. US2003027270A1
GENERAL INFORMATION:
; APPLICANT: Baker,Kevin P.
; APPLICANT: Beresini,Maureen
; APPLICANT: DeForge,Laura
; APPLICANT: Desnoyers,Luc
; APPLICANT: Filvaroff,Ellen
; APPLICANT: Gao,Wei-Qiang
; APPLICANT: Gerritsen,Mary E.
; APPLICANT: Goddard,Audrey
; APPLICANT: Godowski,Paul J.
; APPLICANT: Gurney,Austin L.
; APPLICANT: Sherwood,Steven
; APPLICANT: Smith,Victoria
; APPLICANT: Stewart,Timothy A.
; APPLICANT: Tumas,Daniel
; APPLICANT: Watanabe,Colin K
; APPLICANT: Wood,William
; APPLICANT: Zhang,Zemin
; TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
; TITLE OF INVENTION: ACIDS ENCODING THE SAME
; FILE REFERENCE: P3330R1C353
; CURRENT APPLICATION NUMBER: US/10/175,746
; CURRENT FILING DATE: 2002-06-19
; Prior Application removed - See File Wrapper or Palm
; NUMBER OF SEQ ID NOS: 550
; SEQ ID NO 125
; LENGTH: 4374
; TYPE: DNA
; ORGANISM: Homo Sapien
US-10-175-746-125

Query Match   24.8%;   Score 248;   DB 9;   Length 4374;
Best Local Similarity   76.2%;   Pred. No. 6.2e-64;
Matches   305;   Conservative   0;   Mismatches   95;   Indels   0;   Gaps   0;

```
Qy    451 TCTCAAGAGACAGTCATTGGCTGGGCACGGTGGCTCATGCCTCAAATCCCAGCACCTTGG 510
          ||||| ||||||||  ||| |||  || || ||||||||  ||||||||||||||||| ||
Db   3112 TTTTAAAATATGTACTCTGCCCGGTTGCACTGGGTCACCGCTGTAATCCCAGCACTTTGG 3171

Qy    511 GAGGCTGAGGCAGGAGGATCGACTTGAGGCCAGGAGTTCGAGACCAGCCCTGGCCAATATGG 570
          ||| |||||||||||| ||| ||| ||||||||||||||| ||||||||| |||||||||
Db   3172 CAGGCCGAGATGGGCGGATCTTTTGAGGTCAGGAGTTGAGACCAGCGTTGGCCAACATGG 3231

Qy    571 CGAAACCCTCATTTCTACTAAAAATACAAAAAAACTAACCAGGCGTGGTGGCTTGCCTGTA 630
          |||||| ||||||||  || |||||| ||||||||| ||||||| || |||||| |||||
Db   3232 TGAAACCCGTCTCTACTAAAAATACAAAAAATTAGCCAGGCATGGTGGCCATTTGCCTGTA 3291

Qy    631 ATCCCAGCTACTCGGGAGGCTGAGGCAGGAGAATCGCTTGAACCGGGGAGGCAGAGGTTG 690
          ||||||||||||||||||||||||||||||||||| ||||||||  ||||||||||||||
Db   3292 ATCCCAGCTACTCGGGAGGCTGAGGCAGGAGAATCACTTGAACCTGGGAGGCAGAGGTTG 3351

Qy    691 CAGTGAGCTGAGATCACACCACTGCACTCCAGCCTGGGCGACAGAGGACTCCAGCTT 750
          |||||||||||| ||| ||||||||||||||||||||| |||||| |||| ||| |
Db   3352 CAGTGAGCTGAGATTACACCACTGCACTCCAGCCTGGGTGACAGAGTGAGACTCCATCTC 3411

Qy    751 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAGGAGACCATCACTGCTTCTTTACAGAT 810
          |||||  ||||| || ||| || || ||||| ||  |||||| || | ||  ||||||||
Db   3412 AAAAACTGAAAATAAAAATATTGTATTCCTAACTGAAAATATTTACTTAATCTG 3471

Qy    811 GAAAAAACAGGCTCAGAGGTTGAATCGTTTTTCCTGAAGTC 850
          |||||  | || || ||||||||||||  ||||| |||
Db   3472 GAAAACAAATGTAACTATTTTTTAAAGTGGTTACATCTATTC 3511
```

RESULT 7

US-10-176-918-125
Sequence 125, Application US/10176918
Publication No. US2003027275A1
GENERAL INFORMATION:
; APPLICANT: Baker,Kevin P.
; APPLICANT: Beresini,Maureen
; APPLICANT: DeForge,Laura
; APPLICANT: Desnoyers,Luc
; APPLICANT: Filvaroff,Ellen
; APPLICANT: Gao,Wei-Qiang
; APPLICANT: Gerritsen,Mary E.
; APPLICANT: Goddard,Audrey
; APPLICANT: Godowski,Paul J.
; APPLICANT: Gurney,Austin L.
; APPLICANT: Sherwood,Steven
; APPLICANT: Smith,Victoria
; APPLICANT: Stewart,Timothy A.
; APPLICANT: Tumas,Daniel
; APPLICANT: Watanabe,Colin K
; APPLICANT: Wood,William
; APPLICANT: Zhang,Zemin
; TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
; TITLE OF INVENTION: ACIDS ENCODING THE SAME
; FILE REFERENCE: P3330R1C382
; CURRENT APPLICATION NUMBER: US/10/176,918
; CURRENT FILING DATE: 2002-06-20
; Prior Application removed - See File Wrapper or Palm
; NUMBER OF SEQ ID NOS: 550
; SEQ ID NO 125
; LENGTH: 4374
; TYPE: DNA
; ORGANISM: Homo Sapien
US-10-176-918-125

Query Match   24.8%;   Score 248;   DB 9;   Length 4374;
Best Local Similarity   76.2%;   Pred. No. 6.2e-64;
Matches   305;   Conservative   0;   Mismatches   95;   Indels   0;   Gaps   0;

```
Qy    451 TCTCAAGAGACAGTCATTGGCTGGGCACGGTGGCTCATGCCTCAAATCCCAGCACCTTGG 510
          ||||| ||||||||  ||| |||  || || ||||||||  ||||||||||||||||| ||
Db   3112 TTTTAAAATATGTACTCTGCCCGGTTGCACTGGGTCACCGCTGTAATCCCAGCACTTTGG 3171

Qy    511 GAGGCTGAGGCAGGAGGATCGACTTGAGGCCAGGAGTTCGAGACCAGCCCTGGCCAATATGG 570
          ||| |||||||||||| ||| ||| ||||||||||||||| ||||||||| |||||||||
Db   3172 CAGGCCGAGATGGGCGGATCTTTTGAGGTCAGGAGTTGAGACCAGCGTTGGCCAACATGG 3231

Qy    571 CGAAACCCTCATTTCTACTAAAAATACAAAAAAACTAACCAGGCGTGGTGGCTTGCCTGTA 630
          |||||| ||||||||  || |||||| ||||||||| ||||||| || |||||| |||||
Db   3232 TGAAACCCGTCTCTACTAAAAATACAAAAAATTAGCCAGGCATGGTGGCCATTTGCCTGTA 3291

Qy    631 ATCCCAGCTACTCGGGAGGCTGAGGCAGGAGAATCGCTTGAACCGGGGAGGCAGAGGTTG 690
          ||||||||||||||||||||||||||||||||||| ||||||||  ||||||||||||||
Db   3292 ATCCCAGCTACTCGGGAGGCTGAGGCAGGAGAATCACTTGAACCTGGGAGGCAGAGGTTG 3351

Qy    691 CAGTGAGCTGAGATCACACCACTGCACTCCAGCCTGGGCGACAGAGGACTCCAGCTT 750
          |||||||||||| ||| ||||||||||||||||||||| |||||| |||| ||| |
Db   3352 CAGTGAGCTGAGATTACACCACTGCACTCCAGCCTGGGTGACAGAGTGAGACTCCATCTC 3411

Qy    751 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAGGAGACCATCACTGCTTCTTTACAGAT 810
          |||||  ||||| || ||| || || ||||| ||  |||||| || | ||  ||||||||
Db   3412 AAAAACTGAAAATAAAAATATTGTATTCCTAACTGAAAATATTTACTTAATCTG 3471

Qy    811 GAAAAAACAGGCTCAGAGGTTGAATCGTTTTTCCTGAAGTC 850
          |||||  | || || ||||||||||||  ||||| |||
Db   3472 GAAAACAAATGTAACTATTTTTTAAAGTGGTTACATCTATTC 3511
```

Left column (RESULT 8)

```
Qy   751 AAAAAAAAAAAAAAAAAAAAAAAAAGAGAGACCATCACTGCTGTCCTGCATTCTTACAGAT 810
         ||||| ||||||||||||||||||||  ||||||||||||||||||   ||||
Db  3412 AAAAACTGAAAATAAAAATAAAAAATATGTATTCTCCTAACTGAAAATATTTACTTAATCTG 3471

Qy   811 GAAAAAACAGGCTCAGAGGTTGAATCGTTTTCCTGAAGTC 850
         ||||| |||| |||| ||| | |||| |||||||||||
Db  3472 GAAAACAATGTAACTATTTTTTAAAGTGGTTACATCTATTC 3511
```

RESULT 8
US-10-176-921-125

Sequence 125, Application US/10176921
Publication No. US20030027276A1
GENERAL INFORMATION:
; APPLICANT: Baker,Kevin P.
; APPLICANT: Beresini,Maureen
; APPLICANT: DeForge,Laura
; APPLICANT: Desnoyers,Luc
; APPLICANT: Filvaroff,Ellen
; APPLICANT: Gao,Wei-Qiang
; APPLICANT: Gerritsen,Mary E.
; APPLICANT: Goddard,Audrey
; APPLICANT: Godowski,Paul J.
; APPLICANT: Gurney,Austin L.
; APPLICANT: Sherwood,Steven
; APPLICANT: Smith,Victoria
; APPLICANT: Stewart,Timothy A.
; APPLICANT: Tumas,Daniel
; APPLICANT: Watanabe,Colin K
; APPLICANT: Wood,William
; APPLICANT: Zhang,Zemin
; TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
; TITLE OF INVENTION: ACIDS ENCODING THE SAME
; FILE REFERENCE: P3330R1C288
; CURRENT APPLICATION NUMBER: US/10/176,921
; CURRENT FILING DATE: 2002-06-20
; PRIOR APPLICATION removed - See File Wrapper or Palm
; NUMBER OF SEQ ID NOS: 550
; SEQ ID NO 125
; LENGTH: 4374
; TYPE: DNA
; ORGANISM: Homo Sapien
US-10-176-921-125

```
Query Match       24.8%;   Score 248;   DB 9;   Length 4374;
Best Local Similarity   76.2%;   Pred. No. 6.2e-64;
Matches  305;   Conservative  0;   Mismatches  95;   Indels  0;   Gaps  0;

Qy   451 TCTCAAGAGACAGTCATTGGCTGGGCGACGGTGGCTCATGCCTGCAATCCCAGCACCTTGG 510
         | |||  |  | |   ||||||||| | ||||||||||| |||||||||||||||| |||
Db  3112 TTTTAAAATATGTACTCTGGCCGGTTGCAGTGGCTCACGCCTGTAATCCCAGCACTTTGG 3171

Qy   511 GAGGCTGAGGCAGGCAGGAGGATGAACTTGAGGCCAGGAGTTCGAGACCAGCCTGGCCAATATGG 570
         ||||| ||  | |||| |||| ||| | |||| |||||||||||||||||||| |||| ||||
Db  3172 CAGGCCGAGATGGGCGGGAGATCTTTTGAGGTCAGGAGTTCGAGACCAGCCACTTTGG 3231

Qy   571 CGAAACCTCATTTCTACTAAAAACTAACCAGGCGTGGTGGCTTGTGCCTGTA 630
         |||||||  |  ||||||||||||||||| | |||||||  | | | ||||
Db  3232 TGAAACCCCGTCTCTACTAAAAATAACAAAAATTACCAGGCATGGTGGCATTTGCCCTGTA 3291

Qy   631 ATCCCAGCTACTCGGGAGGCTGAGGCAGGAGAATCGCTTGAACCGGGAGGCGGAGGTTG 690
         ||||||||||| ||||||||||||||||||||| | |||||||||| | |||||||||||
Db  3292 ATCCCAGCTACTCGGGAGGCTGAGGCAGGAGAATCACTTGAACCTGGGAGGCGGAGGTTG 3351

Qy   691 CAGTGAGCTGAGATCACACCACTGCACTCCAGCCTGGGCGACAGAGCGAGACTCCAGCTT 750
         ||| || | ||||  ||||||||||| | ||||||||||||||||||| |||||||||
Db  3352 CAGTGAGCTGAGATTACACCACTGCACTCCAGCCTGGGTGACAGAGTGAGACTCCATCTC 3411

Qy   751 AAAAAAAAAAAAAAAAAAAAAAAAGGAGAGACCATCACTGCTGTCCTGCATTCTTACAGAT 810
         ||||| ||||||||||||||||||  ||||||||||||||||||   ||||
Db  3412 AAAAACTGAAAATAAAAATAAAAAATATGTATTCTCCTAACTGAAAATATTTACTTAATCTG 3471

Qy   811 GAAAAAACAGGCTCAGAGGTTGAATCGTTTTCCTGAAGTC 850
         ||||| |||| |||| ||| | |||| |||||||||||
Db  3472 GAAAACAATGTAACTATTTTTTAAAGTGGTTACATCTATTC 3511
```

Right column (RESULT 9)

```
Qy   811 GAAAAAACAGGCTCAGAGGTTGAATCGTTTTCCTGAAGTC 850
         ||||| |||| |||| ||| | |||| |||||||||||
Db  3472 GAAAACAATGTAACTATTTTTTAAAGTGGTTACATCTATTC 3511
```

RESULT 9
US-10-137-865-125

Sequence 125, Application US/10137865
Publication No. US20030032155A1
GENERAL INFORMATION:
; APPLICANT: Baker,Kevin P.
; APPLICANT: Beresini,Maureen
; APPLICANT: DeForge,Laura
; APPLICANT: Desnoyers,Luc
; APPLICANT: Filvaroff,Ellen
; APPLICANT: Gao,Wei-Qiang
; APPLICANT: Gerritsen,Mary E.
; APPLICANT: Goddard,Audrey
; APPLICANT: Godowski,Paul J.
; APPLICANT: Gurney,Austin L.
; APPLICANT: Sherwood,Steven
; APPLICANT: Smith,Victoria
; APPLICANT: Stewart,Timothy A.
; APPLICANT: Tumas,Daniel
; APPLICANT: Watanabe,Colin K
; APPLICANT: Wood,William
; APPLICANT: Zhang,Zemin
; TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
; TITLE OF INVENTION: ACIDS ENCODING THE SAME
; FILE REFERENCE: P3330R1C154
; CURRENT APPLICATION NUMBER: US/10/137,865
; CURRENT FILING DATE: 2002-05-03
; PRIOR APPLICATION removed - See Palm or File Wrapper
; NUMBER OF SEQ ID NOS: 550
; SEQ ID NO 125
; LENGTH: 4374
; TYPE: DNA
; ORGANISM: Homo Sapien
US-10-137-865-125

```
Query Match       24.8%;   Score 248;   DB 9;   Length 4374;
Best Local Similarity   76.2%;   Pred. No. 6.2e-64;
Matches  305;   Conservative  0;   Mismatches  95;   Indels  0;   Gaps  0;

Qy   451 TCTCAAGAGACAGTCATTGGCTGGGCGACGGTGGCTCATGCCTGCAATCCCAGCACCTTGG 510
         | |||  |  | |   ||||||||| | ||||||||||| |||||||||||||||| |||
Db  3112 TTTTAAAATATGTACTCTGGCCGGTTGCAGTGGCTCACGCCTGTAATCCCAGCACTTTGG 3171

Qy   511 GAGGCTGAGGCAGGCAGGAGGATGAACTTGAGGCCAGGAGTTCGAGACCAGCCTGGCCAATATGG 570
         ||||| ||  | |||| |||| ||| | |||| |||||||||||||||||||| |||| ||||
Db  3172 CAGGCCGAGATGGGCGGGAGATCTTTTGAGGTCAGGAGTTCGAGACCAGCCACTTTGG 3231

Qy   571 CGAAACCTCATTTCTACTAAAAACTAACCAGGCGTGGTGGCTTGTGCCTGTA 630
         |||||||  |  ||||||||||||||||| | |||||||  | | | ||||
Db  3232 TGAAACCCCGTCTCTACTAAAAATAACAAAAATTACCAGGCATGGTGGCATTTGCCCTGTA 3291

Qy   631 ATCCCAGCTACTCGGGAGGCTGAGGCAGGAGAATCGCTTGAACCGGGAGGCGGAGGTTG 690
         ||||||||||| ||||||||||||||||||||| | |||||||||| | |||||||||||
Db  3292 ATCCCAGCTACTCGGGAGGCTGAGGCAGGAGAATCACTTGAACCTGGGAGGCGGAGGTTG 3351

Qy   691 CAGTGAGCTGAGATCACACCACTGCACTCCAGCCTGGGCGACAGAGCGAGACTCCAGCTT 750
         ||| || | ||||  ||||||||||| | ||||||||||||||||||| |||||||||
Db  3352 CAGTGAGCTGAGATTACACCACTGCACTCCAGCCTGGGTGACAGAGTGAGACTCCATCTC 3411

Qy   751 AAAAAAAAAAAAAAAAAAAAAAAAAGGAGAGACCATCACTGCTGTCCTGCATTCTTACAGAT 810
         ||||| ||||||||||||||||||  ||||||||||||||||||   ||||
Db  3412 AAAAACTGAAAATAAAAATAAAAAATATGTATTCTCCTAACTGAAAATATTTACTTAATCTG 3471

Qy   811 GAAAAAACAGGCTCAGAGGTTGAATCGTTTTCCTGAAGTC 850
         ||||| |||| |||| ||| | |||| |||||||||||
Db  3472 GAAAACAATGTAACTATTTTTTAAAGTGGTTACATCTATTC 3511
```

RESULT 10
US-10-140-474-125
Sequence 125, Application US/10140474
Publication No. US20030032156A1
GENERAL INFORMATION:
; APPLICANT: Baker,Kevin P.
; APPLICANT: Beresini,Maureen
; APPLICANT: DeForge,Laura
; APPLICANT: Desnoyers,Luc
; APPLICANT: Filvaroff,Ellen
; APPLICANT: Gao,Wei-Qiang
; APPLICANT: Gerritsen,Mary E.
; APPLICANT: Goddard,Audrey
; APPLICANT: Godowski,Paul J.
; APPLICANT: Gurney,Austin L.
; APPLICANT: Watanabe,Colin K
; APPLICANT: Wood,William
; APPLICANT: Zhang, Zemin
; APPLICANT: Smith,Victoria
; APPLICANT: Stewart,Timothy A.
; APPLICANT: Tumas,Daniel
; APPLICANT: Sherwood,Steven
; TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
; TITLE OF INVENTION: ACIDS ENCODING THE SAME
; FILE REFERENCE: P3330R1C162
; CURRENT APPLICATION NUMBER: US/10/140,474
; CURRENT FILING DATE: 2002-05-06
; PRIOR APPLICATION NUMBER: 2002-05-06
; Prior Application removed - See Palm or File Wrapper
; NUMBER OF SEQ ID NOS: 550
SEQ ID NO 125
; LENGTH: 4374
; TYPE: DNA
; ORGANISM: Homo Sapien
US-10-140-474-125

Query Match   24.8%; Score 248; DB 9; Length 4374;
Best Local Similarity   76.2%; Pred. No. 6.2e-64;
Matches 305; Conservative 0; Mismatches 95; Indels 0; Gaps 0;

Qy   451 TCTCAAGAGACAGTCATTGGCTGGGCACGTGGTCATGCCTGCAATCCAGCACCTTGG 510
           ||||| |||||| | |||||| | |||||||||| ||| |||||||||||||||||||
Db   3112 TTTTAAAATATGTACTCTGGCCGGTTCAGTGGCTCACGCCCTGTAATCCCAGCACTTTGG 3171

Qy   511 GAGGCTGAGGCAGGCAGATGACTTGAGGCCAGGAGTTCGAGACCAGCCTGGCCAATATGG 570
           |||| ||||| || || || ||| ||||||||||||||| | ||||||| ||||||||||
Db   3172 CAGGCCGAGATGGGCGGATCTTTTGAGGTCAGGAGTTGAGACCAGCCTGGCCAACATGG 3231

Qy   571 CGAAACCTCATTTCTACTAAAAATACAAAAAACTAACCAGGCGTGGTGGCTTGTGCCTGTA 630
           ||||| ||| | ||| |||||||||||||||||||||||||||||| ||| |||||||||
Db   3232 TGAAACCCCGTCTCTACTAAAAATACAAAAAATTAGCCAGGCATGGTGGCATTTGCCTGTA 3291

Qy   631 ATCCCAGCTACTCGGGAGGCTGAGGCAGGAGAATCGCTTGAACCCGGGAGGCAGAGGTTG 690
           ||||||||||||| ||||||||||||| |||||| ||||||||| ||||||||||||||
Db   3292 ATCCCAGCTACTCGGGAGGCTGAGGCAGGAGAATCACTTGAACCTGGGAGGCAGAGGTTG 3351

Qy   691 CAGTGAGCTGAGATCACACCACTGCACTCCAGCCTGGGCGACAGAGCGAGACTCCAGCTT 750
           ||||||||||||| || |||||||||||||||||||||||||||| ||| || ||||||
Db   3352 CAGTGAGCTGAGATTACACCACTGCACTCCAGCCTGGGTGACAGAGTGAGACTCCATCTC 3411

Qy   751 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAGGAGACCATCACTGCTTCTTACAGAT 810
           ||||||||||||||||||||||||||||||||||||||| || | | ||| | |||
Db   3412 AAAAACTGAAAATAAAAATAAAAATATTGTATTCTCCTAACTGAAATATTTACTTAATCTG 3471

Qy   811 GAAAAAAACAGGCTCAGAGGTTGAATCGTTTTCCTGAAGTC 850
           ||||| | ||| ||| | | ||| |||||| ||| |||||
Db   3472 GAAAACAATGTAACTATTTTAAAGTGGTTACATCTATTC 3511

RESULT 11
US-10-142-431-125
Sequence 125, Application US/10142431
Publication No. US20030036179A1
; GENERAL INFORMATION:

RESULT 12
US-10-143-114-125
Sequence 125, Application US/10143114
Publication No. US20030036180A1
GENERAL INFORMATION:
; APPLICANT: Baker,Kevin P.
; APPLICANT: Beresini,Maureen
; APPLICANT: DeForge,Laura
; APPLICANT: Desnoyers,Luc
; APPLICANT: Filvaroff,Ellen

; APPLICANT: Baker,Kevin P.
; APPLICANT: Beresini,Maureen
; APPLICANT: DeForge,Laura
; APPLICANT: Desnoyers,Luc
; APPLICANT: Filvaroff,Ellen
; APPLICANT: Gao,Wei-Qiang
; APPLICANT: Gerritsen,Mary E.
; APPLICANT: Goddard,Audrey
; APPLICANT: Godowski,Paul J.
; APPLICANT: Gurney,Austin L.
; APPLICANT: Watanabe,Colin K
; APPLICANT: Wood,William
; APPLICANT: Zhang, Zemin
; APPLICANT: Smith,Victoria
; APPLICANT: Stewart,Timothy A.
; APPLICANT: Tumas,Daniel
; APPLICANT: Sherwood,Steven
; TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
; TITLE OF INVENTION: ACIDS ENCODING THE SAME
; FILE REFERENCE: P3330R1C251
; CURRENT APPLICATION NUMBER: US/10/142,431
; CURRENT FILING DATE: 2002-05-10
; Prior Application removed - See File Wrapper or Palm
; NUMBER OF SEQ ID NOS: 550
SEQ ID NO 125
; LENGTH: 4374
; TYPE: DNA
; ORGANISM: Homo Sapien
US-10-142-431-125

Query Match   24.8%; Score 248; DB 9; Length 4374;
Best Local Similarity   76.2%; Pred. No. 6.2e-64;
Matches 305; Conservative 0; Mismatches 95; Indels 0; Gaps 0;

Qy   451 TCTCAAGAGACAGTCATTGGCTGGGCACGTGGTCATGCCTGCAATCCAGCACCTTGG 510
           ||||| |||||| | |||||| | |||||||||| ||| |||||||||||||||||||
Db   3112 TTTTAAAATATGTACTCTGGCCGGTTCAGTGGCTCACGCCCTGTAATCCCAGCACTTTGG 3171

Qy   511 GAGGCTGAGGCAGGCAGATGACTTGAGGCCAGGAGTTCGAGACCAGCCTGGCCAATATGG 570
           |||| ||||| || || || ||| ||||||||||||||| | ||||||| ||||||||||
Db   3172 CAGGCCGAGATGGGCGGATCTTTTGAGGTCAGGAGTTGAGACCAGCCTGGCCAACATGG 3231

Qy   571 CGAAACCTCATTTCTACTAAAAATACAAAAAACTAACCAGGCGTGGTGGCTTGTGCCTGTA 630
           ||||| ||| | ||| |||||||||||||||||||||||||||||| ||| |||||||||
Db   3232 TGAAACCCCGTCTCTACTAAAAATACAAAAAATTAGCCAGGCATGGTGGCATTTGCCTGTA 3291

Qy   631 ATCCCAGCTACTCGGGAGGCTGAGGCAGGAGAATCGCTTGAACCCGGGAGGCAGAGGTTG 690
           ||||||||||||| ||||||||||||| |||||| ||||||||| ||||||||||||||
Db   3292 ATCCCAGCTACTCGGGAGGCTGAGGCAGGAGAATCACTTGAACCTGGGAGGCAGAGGTTG 3351

Qy   691 CAGTGAGCTGAGATCACACCACTGCACTCCAGCCTGGGCGACAGAGCGAGACTCCAGCTT 750
           ||||||||||||| || |||||||||||||||||||||||||||| ||| || ||||||
Db   3352 CAGTGAGCTGAGATTACACCACTGCACTCCAGCCTGGGTGACAGAGTGAGACTCCATCTC 3411

Qy   751 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAGGAGACCATCACTGCTTCTTACAGAT 810
           ||||||||||||||||||||||||||||||||||||||| || | | ||| | |||
Db   3412 AAAAACTGAAAATAAAAATAAAAATATTGTATTCTCCTAACTGAAATATTTACTTAATCTG 3471

Qy   811 GAAAAAAACAGGCTCAGAGGTTGAATCGTTTTCCTGAAGTC 850
           ||||| | ||| ||| | | ||| |||||| ||| |||||
Db   3472 GAAAACAATGTAACTATTTTAAAGTGGTTACATCTATTC 3511

APPLICANT: Gao,Wei-Qiang
APPLICANT: Gerritsen,Mary E.
APPLICANT: Goddard,Audrey
APPLICANT: Godowski,Paul J.
APPLICANT: Gurney,Austin L.
APPLICANT: Sherwood,Steven
APPLICANT: Smith,Victoria
APPLICANT: Stewart,Timothy A.
APPLICANT: Tumas,Daniel
APPLICANT: Watanabe,Colin K
APPLICANT: Wood,William
APPLICANT: Zhang, Zemin
TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
TITLE OF INVENTION: ACIDS ENCODING THE SAME
FILE REFERENCE: P3330R1C211
CURRENT APPLICATION NUMBER: US/10/143,114
CURRENT FILING DATE: 2002-05-09
Prior Application removed - See Palm or File Wrapper
NUMBER OF SEQ ID NOS: 550

SEQ ID NO 125
LENGTH: 4374
TYPE: DNA
ORGANISM: Homo Sapien
US-10-143-114-125

Query Match   24.8%;   Score 248; DB 9;   Length 4374;
Best Local Similarity   76.2%;   Pred. No. 6.2e-64;
Matches   305; Conservative   0;   Mismatches   95;   Indels   0;   Gaps   0;

Qy 451 TCTCAAGAGACAGTCATTGGCTGGGCACGGTGGCCTCATGCCTGCCAATCCCAGCACCTTGG 510
        |||| ||| || ||||| |||| || |||| |||||||| | |||| |||||||||||||| ||
Db 3112 TTTTAAAATATGTACTCTGGCCGGGGTTGCAGTGGCTCACGCCTGTAATCCCAGCACTTTGG 3171

Qy 511 GAGGCTGAGGCAGGCAGAGAATGACTTGAGGCCAGGAGTTGAGACCAGCCAGCCAATATGG 570
        |||||||||||||||||||| ||| ||| |||| ||| || |||| |||||| ||||||| |||
Db 3172 CAGGCCGAGATGGGGGGATCTTTTGAGGTCAGGAGTTCAGGACCAGCCAGCCAACATGG 3231

Qy 571 CGAAACCTCATTTCTACTAAAAATACAAAAAACTAACCAGGCGTGGTGGCTTGTGCCTGTA 630
        || ||||||||||||||||||||||||||||||||||||||||||||| |||||||||||||
Db 3232 TGAAACCCCGTCTCTACTAAAAATACAAAAAATTAGCCAGGCATGGTGGCCATTTGCCTGTA 3291

Qy 631 ATCCCAGCTACTCGGGAGGCTGAGGCAGGAGAATCGCTTGAACCCGGGAGGCAGAGGTTG 690
        | |||||||||||||||||||||||||| | |||||||| || || |||||||||||||||
Db 3292 ATCCCAGCTACTCGGGAGGCTGAGGCAGGAGAATCACTTGAACCTGGGAGGCGGAGGTTG 3351

Qy 691 CAGTGAGCTGAGATCACACCACTGCACTCCAGCCTGGGCGACAGAGCGAGACTCCAGCTT 750
        ||||||||||||||||||||||||||||||||||||||||||| |||||||||||||||||
Db 3352 CAGTGAGCTGAGATTACACCACTGCACTCCAGCCTGGGTGACAGAGTGAGACTCCATCTC 3411

Qy 751 AAAAAAAAAAAAAAAAAAAAAAAAAAAGGAGACCAATCACTGCTGTCCTGCCATTCTTACAGAT 810
        |||||
Db 3412 AAAAAACTGAAAATAAAAAATAATGTATTCTCCTAACTGAAATATTTACTTAATCTG 3471

Qy 811 GAAAAAAACAGGCTCAGAGAGGTTGAATCGTTTTCCTGAAGTC   850
        ||||| | ||
Db 3472 GAAAACAATGTAACTATTTTAAAGTGGTTACATCTATTC 3511

RESULT 13
US-10-140-002-125
Sequence 125, Application US/10140002
Publication No. US20030037623A1
GENERAL INFORMATION:
APPLICANT: Baker,Kevin P.
APPLICANT: Beresini,Maureen
APPLICANT: DeForge,Laura
APPLICANT: Desnoyers,Luc
APPLICANT: Filvaroff,Ellen
APPLICANT: Gao,Wei-Qiang
APPLICANT: Gerritsen,Mary E.
APPLICANT: Goddard,Audrey
APPLICANT: Godowski,Paul J.
APPLICANT: Gurney,Austin L.

APPLICANT: Sherwood,Steven
APPLICANT: Smith,Victoria
APPLICANT: Stewart,Timothy A.
APPLICANT: Tumas,Daniel
APPLICANT: Watanabe,Colin K
APPLICANT: Wood,William
APPLICANT: Zhang, Zemin
TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
TITLE OF INVENTION: ACIDS ENCODING THE SAME
FILE REFERENCE: P3330R1C59
CURRENT APPLICATION NUMBER: US/10/140,002
CURRENT FILING DATE: 2002-05-06
Prior Application removed - See Palm or File Wrapper
NUMBER OF SEQ ID NOS: 550

SEQ ID NO 125
LENGTH: 4374
TYPE: DNA
ORGANISM: Homo Sapien
US-10-140-002-125

Query Match   24.8%;   Score 248; DB 9;   Length 4374;
Best Local Similarity   76.2%;   Pred. No. 6.2e-64;
Matches   305; Conservative   0;   Mismatches   95;   Indels   0;   Gaps   0;

Qy 451 TCTCAAGAGACAGTCATTGGCTGGGCACGGTGGCCTCATGCCTGCCAATCCCAGCACCTTGG 510
        |||| ||| || ||||| |||| || |||| |||||||| | |||| |||||||||||||| ||
Db 3112 TTTTAAAATATGTACTCTGGCCGGGGTTGCAGTGGCTCACGCCTGTAATCCCAGCACTTTGG 3171

Qy 511 GAGGCTGAGGCAGGCAGAGAATGACTTGAGGCCAGGAGTTGAGACCAGCCAGCCAATATGG 570
        |||||||||||||||||||| ||| ||| |||| ||| || |||| |||||| ||||||| |||
Db 3172 CAGGCCGAGATGGGGGGATCTTTTGAGGTCAGGAGTTCAGGACCAGCCAGCCAACATGG 3231

Qy 571 CGAAACCTCATTTCTACTAAAAATACAAAAAACTAACCAGGCGTGGTGGCTTGTGCCTGTA 630
        || ||||||||||||||||||||||||||||||||||||||||||||| |||||||||||||
Db 3232 TGAAACCCCGTCTCTACTAAAAATACAAAAAATTAGCCAGGCATGGTGGCCATTTGCCTGTA 3291

Qy 631 ATCCCAGCTACTCGGGAGGCTGAGGCAGGAGAATCGCTTGAACCCGGGAGGCAGAGGTTG 690
        | |||||||||||||||||||||||||| | |||||||| || || |||||||||||||||
Db 3292 ATCCCAGCTACTCGGGAGGCTGAGGCAGGAGAATCACTTGAACCTGGGAGGCGGAGGTTG 3351

Qy 691 CAGTGAGCTGAGATCACACCACTGCACTCCAGCCTGGGCGACAGAGCGAGACTCCAGCTT 750
        ||||||||||||||||||||||||||||||||||||||||||| |||||||||||||||||
Db 3352 CAGTGAGCTGAGATTACACCACTGCACTCCAGCCTGGGTGACAGAGTGAGACTCCATCTC 3411

Qy 751 AAAAAAAAAAAAAAAAAAAAAAAAAAAGGAGACCAATCACTGCTGTCCTGCCATTCTTACAGAT 810
        |||||
Db 3412 AAAAAACTGAAAATAAAAAATAATGTATTCTCCTAACTGAAATATTTACTTAATCTG 3471

Qy 811 GAAAAAAACAGGCTCAGAGAGGTTGAATCGTTTTCCTGAAGTC   850
        ||||| | ||
Db 3472 GAAAACAATGTAACTATTTTAAAGTGGTTACATCTATTC 3511

RESULT 14
US-10-142-419-125
Sequence 125, Application US/10142419
Publication No. US20030044945A1
GENERAL INFORMATION:
APPLICANT: Baker,Kevin P.
APPLICANT: Beresini,Maureen
APPLICANT: DeForge,Laura
APPLICANT: Desnoyers,Luc
APPLICANT: Filvaroff,Ellen
APPLICANT: Gao,Wei-Qiang
APPLICANT: Gerritsen,Mary E.
APPLICANT: Goddard,Audrey
APPLICANT: Godowski,Paul J.
APPLICANT: Gurney,Austin L.
APPLICANT: Sherwood,Steven
APPLICANT: Smith,Victoria
APPLICANT: Stewart,Timothy A.
APPLICANT: Tumas,Daniel
APPLICANT: Watanabe,Colin K

```
;  APPLICANT: Wood,William
;  APPLICANT: Zhang,Zemin
;  TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
;  TITLE OF INVENTION:  ACIDS ENCODING THE SAME
;  FILE REFERENCE: P3330R1C244
;  CURRENT APPLICATION NUMBER: US/10/142,419
;  CURRENT FILING DATE: 2002-05-10
;  PRIOR APPLICATION NUMBER: US/10/142,419
;  Prior Application removed - See File Wrapper or Palm
;  NUMBER OF SEQ ID NOS: 550
;  SEQ ID NO 125
;  TYPE: DNA
;  ORGANISM: Homo Sapien
;  LENGTH: 4374
US-10-142-419-125
```

```
Query Match             24.8%;    Score 248;  DB 9;  Length 4374;
Best Local Similarity   76.2%;    Pred. No. 6.2e-64;
Matches  305;  Conservative  0;  Mismatches  95;  Indels  0;  Gaps  0;
```

```
Qy  451 TCTCAAGAGACAGTCAITTGGCTGGGCACGGTGGCTCATGCCTGCAATCCCAGCACCTTGG 510
            | | | | |    |  | | | |    ||||| |  ||| || |||||| |||||| |||
Db 3112 TTTTAAAATATGTACTCCTGGCCGGTTGCAGTGGCTCACGCCTGTAATCCCAGCACTTTGG 3171

Qy  511 GAGGCTGAGGCAGGCAGAGGTGGGTGGCTCATGCCTGCAATCCCAGCACCTTTGG 510
            ||| |  |||||| ||   ||| ||||   |||  || | ||||||||||| ||||
Db 3172 CAGGCCGAGAGTGGGGGGGATCTTTTGAGGTCAGGAGTTTGAGACCAGCACTTTGG 570

Qy  511 GAGGCTGAGGCAGGCAGAATGACTTGAGGCCAGGAGTTCGAGACCAGCCTGGCCAATATGG 570
            |||| |  |||||| ||   ||| ||||   ||| |  || | |||||||||||| ||||
Db 3172 CAGGCCGAGAGTGGGGGGATCTTTTGAGGTCAGGAGTTTGAGACCAGCCTGGCCAACATGG 3231

Qy  571 CGAAACCTCATTTCTACTAAAAATACAAAAAATTAACCAGGCGTGGTGGCTTGTGCCTGTA 630
            ||  ||| |||| | |||||||||||||| ||||| ||| |||| | |||||| |||| |
Db 3232 TGAAACCCGTCTCTACTAAAAATACAAAAAATTAGCCAGGCGTGGTGGCATTGCCTGTA 3291

Qy  631 ATCCCAGCTACTCGGGAGGCTGAGGCAGGAGAATCGGCTTGAACCGGGAGGCGGAGGTTG 690
            |||||| ||||||||||||||||||||||| | | | ||||| || || ||| |||||| 
Db 3292 ATCCCAGCTACTCGGGAGGCTGAGGCAGGAGAATCACTTGAACCTGGGAGGCAGAGGTTG 3351

Qy  691 CAGTGAGCTGAGATCACACCACTGCACTCCAGCCTGGGCGACAGAGCGAGACTCCAGCTT 750
            |||||| ||||  |||||||||||||||||||||||||| ||| | ||||| ||| ||||
Db 3352 CAGTGAGCTGAGATTACACCACTGCACTCCAGCCTGGGTGACAGAGTGAGACTCCATCTC 3411

Qy  751 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAGGAGACCATCACTGCCATTCTTACAGAT 810
            |||||     ||||||||||  | |||||  | ||| |||||| ||| ||||||||||
Db 3412 AAAAACTGAAAATAAAAATATGTATTCTCTAACTGAAAATATTACTTAATCTG 3471

Qy  811 GAAAAAACAGGCTCAGAGGTTGAATCGTTTTCCTGAAGTC 850
            |||||     ||  |||||||||| || | | || || ||
Db 3472 GAAAACAATGTAACTATTTTTAAAGTGGTTACATCTATTC 3511
```

```
RESULT 15
US-10-123-262-125
;  Sequence 125, Application US/10123262
;  Publication No. US20030049816A1
;  GENERAL INFORMATION:
;  APPLICANT: Baker,Kevin P.
;  APPLICANT: Beresini,Maureen
;  APPLICANT: DeForge,Laura
;  APPLICANT: Desnoyers,Luc
;  APPLICANT: Filvaroff,Ellen
;  APPLICANT: Gao,Wei-Qiang
;  APPLICANT: Gerritsen,Mary E.
;  APPLICANT: Goddard,Audrey
;  APPLICANT: Godowski,Paul J.
;  APPLICANT: Gurney,Austin L.
;  APPLICANT: Sherwood,Steven
;  APPLICANT: Smith,Victoria
;  APPLICANT: Stewart,Timothy A.
;  APPLICANT: Tumas,Daniel
;  APPLICANT: Watanabe,Colin K
;  APPLICANT: Wood,William
;  APPLICANT: Zhang, Zemin
;  TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
;  TITLE OF INVENTION:  ACIDS ENCODING THE SAME
;  FILE REFERENCE: P3330R1C38
```

```
;  CURRENT APPLICATION NUMBER: US/10/123,262
;  CURRENT FILING DATE: 2002-04-15
;  Prior Application removed - See File Wrapper or Palm
;  NUMBER OF SEQ ID NOS: 550
;  SEQ ID NO 125
;  TYPE: DNA
;  ORGANISM: Homo Sapien
;  LENGTH: 4374
US-10-123-262-125
```

```
Query Match             24.8%;    Score 248;  DB 9;  Length 4374;
Best Local Similarity   76.2%;    Pred. No. 6.2e-64;
Matches  305;  Conservative  0;  Mismatches  95;  Indels  0;  Gaps  0;
```

```
Qy  451 TCTCAAGAGACAGTCAITTGGCTGGGCACGGTGGCTCATGCCTGCAATCCCAGCACCTTGG 510
            | | | | |    |  | | | |    ||||| |  ||| || |||||| |||||| |||
Db 3112 TTTTAAAATATGTACTCCTGGCCGGTTGCAGTGGCTCACGCCTGTAATCCCAGCACTTTGG 3171

Qy  511 GAGGCTGAGGCAGGCAGAGGTGGGTGGCTCATGCCTGCAATCCCAGCACCTTTGG 510
            ||| |  |||||| ||   ||| ||||   |||  || | ||||||||||| ||||
Db 3172 CAGGCCGAGAGTGGGGGGGATCTTTTGAGGTCAGGAGTTTGAGACCAGCACTTTGG 570

Qy  571 CGAAACCTCATTTCTACTAAAAATACAAAAAATTAACCAGGCGTGGTGGCTTGTGCCTGTA 630
            ||  ||| |||| | |||||||||||||| ||||| ||| |||| | |||||| |||| |
Db 3232 TGAAACCCGTCTCTACTAAAAATACAAAAAATTAGCCAGGCGTGGTGGCATTGCCTGTA 3291

Qy  631 ATCCCAGCTACTCGGGAGGCTGAGGCAGGAGAATCGGCTTGAACCGGGAGGCGGAGGTTG 690
            |||||| ||||||||||||||||||||||| | | | ||||| || || ||| |||||| 
Db 3292 ATCCCAGCTACTCGGGAGGCTGAGGCAGGAGAATCACTTGAACCTGGGAGGCAGAGGTTG 3351

Qy  691 CAGTGAGCTGAGATCACACCACTGCACTCCAGCCTGGGCGACAGAGCGAGACTCCAGCTT 750
            |||||| ||||  |||||||||||||||||||||||||| ||| | ||||| ||| ||||
Db 3352 CAGTGAGCTGAGATTACACCACTGCACTCCAGCCTGGGTGACAGAGTGAGACTCCATCTC 3411

Qy  751 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAGGAGACCATCACTGCCATTCTTACAGAT 810
            |||||     ||||||||||  | |||||  | ||| |||||| ||| ||||||||||
Db 3412 AAAAACTGAAAATAAAAATATGTATTCTCTAACTGAAAATATTACTTAATCTG 3471

Qy  811 GAAAAAACAGGCTCAGAGGTTGAATCGTTTTCCTGAAGTC 850
            |||||     ||  |||||||||| || | | || || ||
Db 3472 GAAAACAATGTAACTATTTTTAAAGTGGTTACATCTATTC 3511
```

```
Search completed: March 30, 2003, 02:17:58
Job time : 258.412 secs
```

THIS PAGE BLANK (USPTO)

OM nucleic - nucleic search, using sw model

GenCore version 5.1.4_p5_4578
Copyright (c) 1993 - 2003 Compugen Ltd.

Run on: March 29, 2003, 13:11:19 ; Search time 5027.09 Seconds
                 (without alignments)
                 9361.127 Million cell updates/sec

Title: US-09-691-219-1
Perfect score: 1617
Sequence: 1 cgcggtgaacgcacgaggg.........aaaaaaaaaaaaaaatgtt 1617
Scoring table: IDENTITY_NUC
                 Gapop 10.0 , Gapext 1.0
Searched: 2054640 seqs, 1455140287B residues

Total number of hits satisfying chosen parameters: 4109280

Post-processing: Minimum Match 0%
Minimum DB seq length: 0   Maximum Match 100%
Maximum DB seq length: 2000000000   Listing first 45 summaries

Database :   GenEmbl:*
                 1: gb_ba:*
                 2: gb_htg:*
                 3: gb_in:*
                 4: gb_om:*
                 5: gb_ov:*
                 6: gb_pat:*
                 7: gb_ph:*
                 8: gb_pl:*
                 9: gb_pri:*
                 10: gb_ro:*
                 11: gb_sts:*
                 12: gb_sy:*
                 13: gb_un:*
                 14: gb_vi:*
                 15: gb_ba:*
                 16: gb_fun:*
                 17: em_hum:*
                 18: em_in:*
                 19: em_mu:*
                 20: em_om:*
                 21: em_or:*
                 22: em_ov:*
                 23: em_pat:*
                 24: em_ph:*
                 25: em_pl:*
                 26: em_ro:*
                 27: em_sts:*
                 28: em_un:*
                 29: em_vi:*
                 30: em_htg_hum:*
                 31: em_htg_inv:*
                 32: em_htg_other:*
                 33: em_htg_mus:*
                 34: em_htg_pln:*
                 35: em_htg_rod:*
                 36: em_htg_mam:*
                 37: em_htg_vrt:*
                 38: em_sy:*
                 39: em_htg_hum:*
                 40: em_htgo_mus:*
                 41: em_htgo_other:*

Pred. No. is the number of results predicted by chance to have a

score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	1617	100.0	1617	6	AX411373	AX411373 Sequence
2	1288.4	79.7	1468	9	BC006251	BC006251 Homo sapi
3	1268	78.1	1280	6	AX405563	AX405563 Sequence
4	813.4	50.3	1818	9	AK094027	AK094027 Homo sapi
5	752.4	46.5	2048	10	BC013478	BC013478 Mus muscu
6	734.4	45.4	1559	10	U50927	U50927 Rattus norv
7	601	37.2	2291	9	AK023491	AK023491 Homo sapi
8	521	32.2	1440	9	AK023504	AK023504 Homo sapi
9	363.6	22.5	11101	6	AX411375	AX411375 Sequence
10	363.6	22.5	176006	9	AL391650	AL391650 Human DNA
11	350.6	21.7	1952	10	MMU7007	U76007 Mus musculu
12	346	21.4	2000	9	HSU76010	U76010 Human putat
13	346	21.4	2101	9	BC028358	BC028358 Homo sapi
14	309.6	19.1	2823	6	AX061210	AX061210 Sequence
15	309.6	19.1	2852	6	AX351593	AX351593 Sequence
16	306.4	18.9	5422	9	AX086187	AX086187 Sequence
17	306.4	18.9	5422	9	HSM800077	HSM800077 Homo sapi
18	277.6	17.2	153241	9	AC109690	AC109690 Homo sapi
19	269.2	16.6	198220	9	AC102314	AC102314 Mus muscu
20	264.4	16.4	696	9	AC102314	AC102314 Mus muscu
21	262.6	16.2	761	9	HSA337344	HSA337344 Homo sapi
22	251.8	15.6	743	9	HSA336978	HSA336978 Homo sapi
23	197.4	12.2	719	9	AJ336978	AJ336978 Homo sapi
24	184.2	11.4	5276	10	AF003747	AF003747 Mus muscu
25	177.2	11.0	1720	6	BC026089	BC026089 Homo sapi
26	176.2	10.9	5420	10	RNY16774	Y16774 Rattus norv
27	175.6	10.9	1557	9	AF025409	AF025409 Homo sapi
28	167	10.3	1751	3	AY071460	AY071460 Drosophil
29	152.6	9.4	26209	3	CEY39E4A	CEY39E4A Caenorhab
30	152.6	9.4	81100	3	AC019838	AC019838 Drosophil
31	152.6	9.4	171134	3	AC009198	AC009198 Drosophil
32	152.6	9.4	174314	3	AC011066	AC011066 Drosophil
33	152.6	9.4	307911	3	AE003635	AE003635 Drosophil
34	151.6	9.4	1083	8	AY044454	AY044454 Thlaspi g
35	151	9.3	110000	2	CEY111B2_1	CEY111B2_1
36	151	9.1	72421	2	AC101509	AC101509 Mus muscu
37	146.6	9.1	138537	2	AC130608	AC130608 Oryza sat
38	132	8.2	143506	2	AC130598	AC130598 Oryza sat
39	131.2	8.1	800	8	HSA331528	HSA331528 Homo sapi
40	126.6	7.8	1179	8	AY044452	AY044452 Thlaspi g
41	125.6	7.8	160040	2	AC105674	AC105674 Rattus no
42	125.6	7.7	174303	2	AC099104	AC099104 Rattus no
43	124.8	7.7	240125	10	AL669982	AL669982 Mouse DNA
44	121.8	7.5	1179	8	AY044453	AY044453 Thlaspi g
45	119.8	7.4	142001	8	AL138642	AL138642 Arabidops

ALIGNMENTS

RESULT 1
AX411373

LOCUS   AX411373   1617 bp   DNA   linear   PAT 14-JUN-2002
DEFINITION   Sequence 1 from Patent WO0224910.
ACCESSION   AX411373
VERSION   AX411373.1 GI:21444024
KEYWORDS   .
SOURCE   human.
  ORGANISM   Homo sapiens
             Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
             Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1
  AUTHORS   Wei,M.H., Ketchum,K.A., di Francesco,V. and Beasley,E.M.
  TITLE   Isolated human transporter proteins, nucleic acid molecules
             encoding them, and uses thereof

```
JOURNAL      Patent: WO 0224910-A 1 28-MAR-2002;
             PE Corporation (NY) (US)
FEATURES          Location/Qualifiers
  source      1..1617
             /organism="Homo sapiens"
             /db_xref="taxon:9606"
BASE COUNT      351 a      473 c      489 g      304 t
ORIGIN

Query Match      100.0%;     Score 1617;    DB 6;   Length 1617;
Best Local Similarity   100.0%;   Pred. No. 0;
Matches 1617;   Conservative  0;   Mismatches  0;   Indels  0;   Gaps  0;

Qy      1 CGCGGGTGAACAAGCACGCAGGGGGGAGCGGCCGCTTGGCACGGCGGGAGCTGCAGCC 60
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1 CGCGGGTGAACAAGCACGCAGGGGGGAGCGGCCGCTTGGCACGGCGGGAGCTGCAGCC 60

Qy     61 GCAGCGGGGAGACCGGGGGAGCTCAGGGGGGCCCCAGGAGGTGAGCGCACCTCG 120
          |||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     61 GCAGCGGGGAGACCGGGGGAGCTCAGGGGGGCCCCAGGAGGTGAGCGCACCTCG 120

Qy    121 CGGGGCCCTCGGGGAGGTGGGGTGAGGCGCCCCACCCCGAGCCCAACTTTCAGGGC 180
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    121 CGGGGCCCTCGGGGAGGTGGGGTGAGGCGCCCCACCCCGAGCCCAACTTTCAGGGC 180

Qy    181 GCACTCGGGGCGGGGCGGCTGCCGCGGGCTCGGGGCCGGGGGGCTCGGGGAGGCCAA 240
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    181 GCACTCGGGGCGGGGCGGCTGCCGCGGGCTCGGGGCCGGGGGGCTCGGGGAGGCCAA 240

Qy    241 GGAGAAGCACGACGATCTGTTGGACGCGCGGGGCAATCCGGTCATACACGGGATCTCTGTG 300
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    241 GGAGAAGCACGACGATCTGTTGGACGCGCGGGGCAATCCGGTCATACACGGGATCTCTGTG 300

Qy    301 GCAGGAAGGGGGCTGGCTGGGATTCCTCTGCCCCGACCTGGCTTGCAGGGCCAATGA 360
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    301 GCAGGAAGGGGGCTGGCTGGGATTCCTCTGCCCCGACCTGGCTTGCAGGGCCAATGA 360

Qy    361 GCTGGCTGCCCGAGAGCAAACATCACTGCATGCCTCAGAGAAGTGGTCCTTGCACTGTTGA 420
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    361 GCTGGCTGCCCGAGAGCAAACATCACTGCATGCCTCAGAGAAGTGGTCCTTGCACTGTTGA 420

Qy    421 CCCCAAGAAAGGGGAAGGGGAAGCCCCAGCGCCGCAGCGTGTATGTAGCCTCTGCCTGTTGTT 480
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    421 CCCCAAGAAAGGGGAAGGGGAAGCCCCAGCGCCGCAGCGTGTATGTAGCCTCTGCCTGTTGTT 480

Qy    481 CATGATCGGAGAGAAGTCGTTGGTGGGGTACCAGTCATGACTGACGC 540
          ||||||||||||||||||||||||||||||||||||||||||||||
Db    481 CATGATCGGAGAGAAGTCGTTGGTGGGGTACCAGTCATGACTGACGC 540

Qy    541 AGCACACCTGCTCACTGACTTTGCCAGCATGCTCATCAGCCCTCTTCTCCCCTGGATGTC 600
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    541 AGCACACCTGCTCACTGACTTTGCCAGCATGCTCATCAGCCCTCTTCTCCCCTGGATGTC 600

Qy    601 CTCCCCGCCCAGCCACCAAGACCATGAACTTTGGCTGCCTGGGCAGGAGCTTGGGCAGGTGGAGC 660
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    601 CTCCCCGCCCAGCCACCAAGACCATGAACTTTGGCTGCCTGGGCAGGAGCTTGGGCAGGTGGAGC 660

Qy    661 CCTGGTCTCTGTACTGTCCATCTGGGTCGGGGGTACCTGGCTCTGGGA 720
          ||||||||||||||||||||||||||||||||||||||||||||||
Db    661 CCTGGTCTCTGTACTGTCCATCTGGGTCGGGGGTACCTGGCTCTGGGA 720

Qy    721 GCGGCTGATCTCTGGGACTATGAAAATTGACGGGGGGACCATGCTCGGGCTG 780
          ||||||||||||||||||||||||||||||||||||||||||||||||
Db    721 GCGGCTGATCTCTGGGACTATGAAAATTGACGGGGGGACCATGCTCGGGCTG 780

Qy    781 CGCTCGTGGCGTGTGAACATCATAATGGGGTTGACCCTTCACCAGGTTCACGTT 840
          |||||||||||||||||||||||||||||||||||||||||||||||||
Db    781 CGCTCGTGGCGTGTGAACATCATAATGGGGTTGACCCTTCACCAGGTTCACGTT 840

Qy    841 CCACGGCACCACCAACCAGGAGGAGGAGGAACCCCAGCGTCCGAGCTGCCTTCATCCATGT 900
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    841 CCACGGCACCACCAACCAGGAGGAGGAGGAACCCCAGCGTCCGAGCTGCCTTCATCCATGT 900
```

```
Qy    901 GATCGGGGCCGACTTTATGCAGAGCATGGGGTGTCCTTGGGGTGCCAGCGCCTATATTTTATTACTTCAA 960
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    901 GATCGGGGCCGACTTTATGCAGAGCATGGGGTGTCCTTGGGGTGCCAGCGCCTATATTTTATTACTTCAA 960

Qy    961 GCCAGAATACAAGTAGTAGACCCCATCCTCACCTTCGTCCTTCCATCCTGGTCCTTGGG 1020
          |||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    961 GCCAGAATACAAGTAGTAGACCCCATCCTCACCTTCGTCCTTCCATCCTGGTCCTTGGG 1020

Qy   1021 GACAACCTTGACCATCCTCGAGAGAGTGTGATCCTCTTGGGTTTGAGTGGGAAGGACCCCCAAGG 1080
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1021 GACAACCTTGACCATCCTCGAGAGAGTGTGATCCTCTTGGGTTTGAGTGGGAAGGACCCCCAAGG 1080

Qy   1081 CGTTGACTTCACAGCTGTTCGTGCGTCATCCTGCACCTTCGTCGGTGGGAGGGGGTTAGAAGCCCTGCA 1140
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1081 CGTTGACTTCACAGCTGTTCGTGCGTCATCCTGCACCTTCGTCGGTGGGAGGGGGTTAGAAGCCCTGCA 1140

Qy   1141 CAGCCTGCATATCTGGGCAACTGACGGTGCCCAGCCAGCCTGTTCTGTCCGGTCGGTGCCCAT 1200
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1141 CAGCCTGCATATCTGGGCAACTGACGGTGCCCAGCCAGCCTGTTCTGTCCGGTCGGTGCCCAT 1200

Qy   1201 TGCTCAGAATACAGAGACCCCAGCGCTGTCTGAAGACAACGAGCCAGCACCCTGTTGCCTCAAGGGAA 1260
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1201 TGCTCAGAATACAGAGACCCCAGCGCTGTCTGAAGACAACGAGCCAGCACCCTGTTGCCTCAAGGGAA 1260

Qy   1261 GTTCCACTTCCACACCCGTGACCATCCAGATCGAAGGACTACTCGGAGGACATGAAGGACTG 1320
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1261 GTTCCACTTCCACACCCGTGACCATCCAGATCGAAGGACTACTCGGAGGACATGAAGGACTG 1320

Qy   1321 TCAGGCAATGCCAGGGCCCCTCAGACTGACTGACTGCTCAGCGACTGCTTCAGCAGTCACTGTGA 1380
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1321 TCAGGCAATGCCAGGGCCCCTCAGACTGACTGACTGCTCAGCGACTGCTTCAGCAGTCACTGTGA 1380

Qy   1381 AGGACCTGCAGTGTGGGCTGGGACTGAGTGTCTCCCCCAGGGACTGACTTTGCCTCCTACCCC 1440
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1381 AGGACCTGCAGTGTGGGCTGGGACTGAGTGTCTCCCCCAGGGACTGACTTTGCCTCCTACCCC 1440

Qy   1441 AGCTGTGTTATAAAACCAGGTCCCCCTCCTGACCTCTGCCCCCACTCCAGGAAATGGGAGCTCT 1500
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1441 AGCTGTGTTATAAAACCAGGTCCCCCTCCTGACCTCTGCCCCCACTCCAGGAAATGGGAGCTCT 1500

Qy   1501 TCCCAGCCTCCCATCTGACTACAGCCCAGGTGGGGACTCAGCGGCTATAAAGCTAGTGTG 1560
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1501 TCCCAGCCTCCCATCTGACTACAGCCCAGGTGGGGACTCAGCGGCTATAAAGCTAGTGTG 1560

Qy   1561 ACCCTGAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAATGTT 1617
          ||||||||||||||||||||||||||||||||||||||||||||||
Db   1561 ACCCTGAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAATGTT 1617
```

```
RESULT 2
LOCUS       BC006251
DEFINITION  BC006251                    1468 bp     mRNA     linear    PRI 12-JUL-2001
            Homo sapiens, Similar to zink transporter 2, clone MGC:11303
            IMAGE:3948165, mRNA, complete cds.
ACCESSION   BC006251
VERSION     BC006251.1     GI:13623300
KEYWORDS    MGC.
SOURCE      Homo sapiens
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 1468)
  AUTHORS   Strausberg,R.
  TITLE     Direct Submission
  JOURNAL   Submitted (09-APR-2001) National Institutes of Health, Mammalian
            Gene Collection (MGC), Cancer Genomics Office, National Cancer
            Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,
            USA
REMARK      NIH-MGC Project URL: http://mgc.nci.nih.gov
COMMENT     Contact: MGC help desk
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: DCTD/DPP
            cDNA Library Preparation: Rubin Laboratory
```

```
            cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
            DNA Sequencing by: National Institutes of Health Intramural
            Sequencing Center (NISC),
            Gaithersburg, Maryland;
            Web site:    http://www.nisc.nih.gov/
            Contact:     nisc_mgc@nhgri.nih.gov
            Shevchenko,Y., Wetherby,K.D., Beckstrom-Sternberg,S.M.,
            Benjamin,B., Blakesley,R.W., Bouffard,G.G., Brinkley,C., Brooks,S.,
            Dietrich,N.L., Guan,X., Gupta,J., Ho,S.-L., Karlins,E., Legaspi,R.,
            Lim,M., Maduro,Q.L., Masiello,C., Mastrian,S.D., McCloskey,J.C.,
            McDowell,J., Pearson,R., Snyder,E., Stantripop,S., Thomas,P.J.,
            Tiongson,E.E., Touchman,J.W., Tsurgeon,C., Vogt,J.L., Walker,M.A.,
            Zhang,L.-H. and Green,E.D.

            Clone distribution: MGC clone distribution information can be found
            through the I.M.A.G.E. Consortium/LLNL at: http://image.llnl.gov
            Series: IRAL Plate: 15 Row: h Column: 11
            This clone was selected for full length sequencing because it
            passed the following selection criteria: Hexamer frequency ORF
            analysis.

FEATURES             Location/Qualifiers
     source          1..1468
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="MGC:11303 IMAGE:3948165"
                     /tissue_type="Ovary, adenocarcinoma"
                     /clone_lib="NIH_MGC_9"
                     /lab_host="DH10B-R"
                     /note="Vector: pOTB7"
     CDS             227..1198
                     /codon_start=1
                     /product="Similar to zinc transporter 2"
                     /protein_id="AAH06251.1"
                     /db_xref="GI:13623301"
                     /translation="MEAKEKQHLLDARPAIRSYTGSLWQEGAGWIPLPRPGLDLQAIE
                     LAAQSNHHCHAQKGPDSHCDPKKGKRAQROLYVASAICLLFMIGEVVEILGALVSVLSI
                     WVVIGVLVYLAVERLISGDYEIDGGTMLITSGCAVAVNIIMGLILHQSGHGHSHGTTN
                     QOEENPSVRAAFIHVIGDPMQSMGVLVAAYILYFKPEYKYVDPICTFVFSILVLGTTL
                     TILRDVILVLMEGTPKGVDFTAVRDLLLSVEGVEALHSLHIWALTVAQPVLSVHITAIA
                     QNTDAQAVLKTASSRLQGKFHPHTVTIQIEDYSEDMKDCQACQGPSD"

BASE COUNT       327 a     423 c     451 g     267 t
ORIGIN

Query Match       79.7%;  Score 1288.4;  DB 9;  Length 1468;
Best Local Similarity   90.7%;  Pred. No.1.6e-265;
Matches 1446;  Conservative   0;  Mismatches   1;  Indels   147;  Gaps   1;

Qy     20 GGGAGCGGCTTGGCACCCGGACCCAGAGCCGAGCTCCAGCCGCCAGCGGGGACCCGGGGGAG   79
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     17 GGGGAGCGGCTTGGCACGCGGACCCAGAGCCGAGCTCCAGCCGCCAGCGGGGACCCGGGGGAG   76

Qy     80 CTCAGGGACCCGCAGGAGAGCGGGGCGGGGGCGCACCTCGGCGGGGCCCCTCGGCGCCAGGT  139
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     77 CTCAGGGACCCGCAGAGAGCGGGGCGGGGGCGCACCTCGGCGGGGCCCCTCGGCGCCAGGT  136

Qy    140 GGGTGAGCGGCCACCCGGAGTCCCGCGCGCCAACTTTCAGGGGGCCCACTCGGGGGCGGCTG  199
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    137 GGGTGAGCGGCCACCCGGAGTCCCGCGCGCCAACTTTCAGGGGGCCCACTCGGGGGCGGCTG  196

Qy    200 CGCGGCTGCCGGGACTCGCGGCCGCGGGAGAGGAGCACCACATCTGTTG  259
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    197 CGCGGCTGCCGGGACTCGCGGCCGCGGGAGAGGAGCACCACATCTGTTG  256

Qy    260 GACGCCAGGCCGGCAATCCGGTCATCACACGGGATCTCTGTGGCCAGGAAAGGGGCTGGCTGG  319
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    257 GACGCCAGGCCGGCAATCCGGTCATCACACGGGATCTCTGTGGCCAGGAAAGGGGCTGGCTGG  316

Qy    320 ATTCGCGTCTGCCCGACCTGGCCTGGGCTGGCCGACTTGCAGGCCGCCAGGCCAAC  379
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    317 ATTCCTCTCGCCCGACCTGGCCTGGGCTGGCCATTGAGCTGGCCTGCCCAGAGCAAC  376

Qy    380 CATCACTGCCATGCTCAGAAGGGTCCTGACAGTCACTGTGACCCCAAGAAGGGAAGGCC  439
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    380 CATCACTGCCATGCTCAGAAGGGTCCTGACAGTCACTGTGACCCCAAGAAGGGAAGGCC  436

Db    377 CATCACTGCCATGCTCAGAAGGGTCCTGACAGTCACTGTGACCCCAAGAAGGGAAGGCC  436

Qy    440 CACGCGCCAGCCTGTATGTAGCCCTCTGCCAATCTCGCCTGTTGTTTCATGATCGGAAGTCGTT  499
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    437 CACGCGCCAGCGTGTATGTAGCCCTCTGCCAATCTCGCCTGTTGTTTCATGATCGGAAGTCGT-  495

Qy    500 GGTGGGTACCTGGCACACAGCTTGGCGTGCATGACTGACGCAGCACACTGCTCACTGAC  559
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    496 ------------------------------------------------------------  495

Qy    560 TTTGGCCACATGCTCATCAGCCTCTTCTCCCCTCTCGGATGTCCTCCCGGCCAGCCACCAAG  619
          ------------------------------------------------------------
Db    496 ------------------------------------------------------------  495

Qy    620 ACCATGAACTTTGGCTGCCAGAGAGCTCGAGATCTTGGCTCGTTGTTCTTGTACTGTCC  679
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    496 --------------TGAGACTTGGGAGCCCTGGGTCTCTGTACTGTCC  529

Qy    680 ATCTGGGTGCTGCGTGACGGGGGGGTACTGGTGTTGTACCTGGGCTGTTGGACGGGGGGAC  739
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    530 ATCTGGGTGCTGCGTGACGGGGGGGTACTGGTGTTGTACCTGGGCTGTTGGACGGGGGGAC  589

Qy    740 TATGGAAAATTGACGGGGGGGACCCATGCTGATCAGGTCGGGCGTGCGGCGTGTGGAACATC  799
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    590 TATGGAAAATTGACGGGGGGGACCCATGCTGATCAGGTCGGGCGTGCGGCGTGTGGAACATC  649

Qy    800 ATAAATGGGTTGACCCTTCACCAGTCTGGCCATGGGCCACAGCCCACCACCAACCAG  859
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    650 ATAAATGGGTTGACCCTTCACCAGTCTGGCCATGGGCCACAGCCCACCACCAACCAG  709

Qy    860 CAGGAGGAGAACCCCAGCGTCCGAGCGTCCTTCATCCATGTGATCGGGGGACTTTATGCAG  919
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    710 CAGGAGGAGAACCCCAGCGTCCGAGCTGCCTTCATCCATGTGATCGGGGGACTTTATGCAG  769

Qy    980 GACCCCATCTGCACCTTCGTCTTCTCCATCCTGGTCCTGGGGGACAACCTTGACCATCTG  1039
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    770 AGCATGGGGGTGCTCGTGGCAGCCTATATTTTGTATTTCAAGCCAGAATACAAGTATGTA  829

Qy    920 AGCATGGGGGTGCTCGTGGCAGCCTATATTTTGTATTTCAAGCCAGAATACAAGTATGTA  979
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    830 GACCCCATCTGCACCTTCGTCTTCTCCATCCTGGTCCTGGGGGACAACCTTGACCATCTG  889

Qy   1040 AGAGATGTGATCCTGGTGTTGATGGAAGGGACCCCCAAGGGGGCGTTGACTTCACGCTGTT  1099
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    890 AGAGATGTGATCCTGGTGTTGATGGAAGGGACCCCCAAGGGGGCGTTGACTTCACAGCTGTT  949

Qy   1100 CTTGATCTGTCTGCTCTGTCTGGTGGAGGGGTAGAAGGAGCCCCTGCACAGCCCCATATCTGGGCA  1159
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    950 CGTGATCTGCTGCTCTGTCTGCGGTGGAGGGGTAGAAGCCCCTGCACAGCCCCATATCTGGGCA  1009

Qy   1160 CTGACGCGGTGGCCCCAGCCTGTTCTGTCTCTGTCACACTCGCCAATGCATCAGAATACAGAGAGCC  1219
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1010 CTGACGCGGTGGCCCCAGCCTGTTCTGTCTCTGTCACATCGCCAATGCATCAGAATACAGAGAGCC  1069

Qy   1220 CAGGGCCTGTGCTGAAGACAGCAGCAGCAGCGCCGCCTCCAAGGGGGAAGTTCCACTTCACACCGTG  1279
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1070 CAGGCTGTGGCTGAAGACAGCAGCAGCAGCGCCGCCTCCAAGGGGGAAGTTCCACTTCACACCGTG  1129

Qy   1280 ACCATCCAGATCGAGGAGGACTACTCGGAGAGGACATGAAGGACATGAAAGGAGCCCC  1339
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1130 ACCATCCAGATCGAGGAGGACTACTCGGAGAGGACATGAAGGACATGAAAGGAGCCCC  1189

Qy   1340 TCAGACTGACTCGCTCAGCGACGCCAACTGGGCCATGAACAGGACCTGGCAGTTGGCTGG  1399
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1190 TCAGACTGACTGCTCAGCGACGCCAACTGGGCCATGAACAGGACCTGGCAGTTGGCTGG  1249

Qy   1400 ACTGAGGTGTCCCCCAGGCCCAGCCCAGGACTTTGCCTACCCCACCAGCTGTGTTTATAAACCAGG  1459
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1250 ACTGAGGTGTCCCCCAGGCCCAGCCCAGGACTTTGCCTACCCCACCAGCTGTGTTTATAAACCAGG  1309

Qy   1460 TCCCCCTCCTGACCTCTGCCCCACTCCAGGAATGGAGGCTCTTCCCCAGGCCTCCCCATCTGAC  1519
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1310 TCCCCCTCCTGACCTCTGCCCCACTCCAGGAATGGAGGCTCTTCCCCAGGCCTCCCCATCTGAC  1369
```

RESULT 3

LOCUS AX405563 1280 bp DNA linear PAT 14-JUN-2002
DEFINITION Sequence 30 from Patent WO0222684.
ACCESSION AX405563
VERSION AX405563.1 GI:21438588
KEYWORDS
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1
AUTHORS Lee,E.A., Yue,H., Lal,P.G., Walia,N.K., Baughn,M.R., Warren,B.A.,
Lee,S., Sanjawala,M.S., Yao,M.G., Ramkumar,J., Thornton,M.,
Gandhi,A.R., Policky,J.L., Elliott,V.S., Arvizu,C., Raumann,B.E.,
Bruns,C.M., Naini,A., Hafalia,A.J., Nguyen,D.B., Xu,Y., Lu,D.A.,
Ison,C.H., Griffin,J.A., Reddy,R.M. and Burford,N.
TITLE Transporters and ion channels
JOURNAL Patent: WO 0222684-A 30 21-MAR-2002;
Incyte Genomics, Inc. (US)
FEATURES Location/Qualifiers
source 1..1280
/organism="Homo sapiens"
/db_xref="taxon:9606"
/note="Incyte ID No: 2614283CB1"
BASE COUNT 248 a 383 c 392 g 257 t
ORIGIN

Query Match 78.4%; Score 1268; DB 6; Length 1280;
Best Local Similarity 99.9%; Pred. No. 3.7e-261;
Matches 1279; Conservative 0; Mismatches 0; Indels 1; Gaps 1;

Qy 88 CCGCAGGAGGCCGGGAGTGAGGCACCTCGGGGGG-CCCTCGGGGAGGTGGGTGAG 146
Db 1 CCGCAGGAGGCCGGGAGTGAGGCACCTCGGGGGGCCCCTCGGGGAGGTGGGTGAG 60

Qy 147 CGCCACCCGGAGTCCCGCCGCCGCAACTTTCAGGGGGCGGGGCGGGCTGCCGGGCT 206
Db 61 CGCCACCCGGAGTCCCGCCGCCGCAACTTTCAGGGGGCGGGGCGGGCTGCCGGGCT 120

Qy 207 GCGCGGGACTCGGGCGCGGGACTGCATGGGGAGGCCAAGGAGAGAGAGAAAGCAGACATCTGTTGGACGCCA 266
Db 121 GCGCGGGACTCGGGCGCGGGACTGCATGGGGAGGCCAAGGAGAGAGAGAAAGCAGACATCTGTTGGACGCCA 180

Qy 267 GGGCGGGCAATCCGGTCATACACGCGGGGATCTTCTGTGGCCAGGAAGGGGCTGGCTGGATTCCTC 326
Db 181 GGGCGGGCAATCCGGTCATACACGCGGGGATCTTCTGTGGCCAGGAAGGGGCTGGCTGGATTCCTC 240

Qy 327 TGGCCCCGACCTGGCCCTGGGACTTGCAGGCCCATTGAGCTGCCCAGCGCAACCATCACT 386
Db 241 TGCCCCGACCTGGCCCTGGGACTTGCAGGCCCATTGAGCTGCCCAGCGCAACCATCACT 300

Qy 387 GCCATGCTCAGAAGGGTCCTGACAGTCACTGTGACCCCAAGAGAGGGGAAGGGCCCAGCCGCC 446
Db 301 GCCATGCTCAGAAGGGTCCTGACAGTCACTGTGACCCCAAGAGAGGGGAAGGGCCCAGCCGCC 360

Qy 447 AGCTGTATGTAGTAGCCTCTGCCATCTGCCTGTTGTTCATGATGATCGGAGAAGTCGTTGGTGGGT 506
Db 361 AGCTGTATGTAGTAGCCTCTGCCATCTGCCTGTTGTTCATGATGATCGGAGAAGTCGTTGGTGGGT 420

Qy 507 ACCTGGCACACAGCTTGGCTGTGTCATGACTGACCGACGCACAACCTGACTTTGCCA 566
Db 421 ACCTGGCACACAGCTTGGCTGTGTCATGACTGACCGACGCACAACCTGACTTTGCCA 480

RESULT 4

LOCUS AK094027 1818 bp mRNA linear PRI 15-JUL-2002
DEFINITION Homo sapiens cDNA FLJ36708 fis, clone UTERU2009904,
similar to ZINC TRANSPORTER 2.
ACCESSION AK094027
VERSION AK094027.1 GI:21753005
KEYWORDS oligo capping; fis (full insert sequence).
SOURCE Homo sapiens uterus cDNA to mRNA, clone_lib:UTERU2
clone:UTERU2009904.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1
AUTHORS Suzuki,O., Sasaki,N., Aotsuka,S., Shoji,T., Ichihara,T.,

Qy 567 GCATGCTCATCAGCCTCTTCTCCCCTCTCGGATCTGCTCCCGGCTCCACCCAAGAGACCATGA 626
Db 481 GCATGCTCATCAGCCTCTTCTCCCCTCTCGGATCTGCTCCCGGCTCCACCCAAGAGACCATGA 540

Qy 627 ACTTTGGCTGGCAGAGACTGAGATCTTGGGAGCCCTGGCTCTTGTGTCCATCTGGG 686
Db 541 ACTTTGGCTGGCAGAGACTGAGATCTTGGGAGCCCTGGCTCTTGTGTCCATCTGGG 600

Qy 687 GGTTGACGGGGCGTACTGGTGTGACCTGCGTGGAGGGCGATCCTGGGGGACTATGAA 746
Db 601 GGTTGACGGGGCGTACTGGTGTGACCTGCGTGGAGGGCGATCCTGGGGGACTATGAA 660

Qy 747 TTGACCGGGGGGACCATGCTGATCACGTGGCGGGTGGCTGCCTGTGGTGAACATCATAATGG 806
Db 661 TTGACCGGGGGGACCATGCTGATCACGTGGCGGGTGGCTGCCTGTGGTGAACATCATAATGG 720

Qy 807 GGTTGACCCTTCACCCAGTCTGGCCATGGGCACACAGCCACCAGGCACCAAACCAGCAGGAGG 866
Db 721 GGTTGACCCTTCACCCAGTCTGGCCATGGGCACACAGCCACCAGGCACCAAACCAGCAGGAGG 780

Qy 867 AGAAACCCCAGGGGTCCGAGCTGCCTTCATCCCARTGTGAGCCGACGTTTTATGCAGAGCATGG 926
Db 781 AGAAACCCCAGGGGTCCGAGCTGCCTTCATCCCARTGTGAGCCGACGTTTTATGCAGAGCATGG 840

Qy 927 GTGTCCTAGTGCAGCCTATATTTTAACTTCAAGCCAGAATACAAGTATGTAGAGCCCCA 986
Db 841 GTGTCCTAGTGCAGCCTATATTTTAACTTCAAGCCAGAATACAAGTATGTAGAGCCCCA 900

Qy 987 TCTGCACCTTCGTCTTCTCCCATCCTCGGTCCTGCGGGACAAACCTTGACCAATCCTGAGAGATG 1046
Db 901 TCTGCACCTTCGTCTTCTCCCATCCTCGGTCCTGCGGGACAAACCTTGACCAATCCTGAGAGATG 960

Qy 1047 TGATCCTGGTGTTGATGGAAGGGACCCCCAGGAGGCCGTTGACTTCACAGGCTGTTTCGTGGATC 1106
Db 961 TGATCCTGGTGTTGATGGAAGGGACCCCCAGGAGGCCGTTGACTTCACAGGCTGTTTCGTGGATC 1020

Qy 1107 TGCTGCTGTCTCGGTGGACGGGGTTAGAAGCCCTGCCACAGCCTCCATATCTGGGCACTGACGG 1166
Db 1021 TGCTGCTGTCTCGGTGGACGGGGTTAGAAGCCCTGCCACAGCCTCCATATCTGGGCACTGACGG 1080

Qy 1167 TGGCCCAGCCTGTTCTGTTCTGTTCCACATCGCCATGCCCTGCTCAGAATACAGACCCCCAGCCTG 1226
Db 1081 TGGCCCAGCCTGTTCTGTTCTGTTCCACATCGCCATGCCCTGCTCAGAATACAGACCCCCAGCCTG 1140

Qy 1227 TGCTGAAGACAGCCACGCCAGCACCGCGCCCTCCAAGGGAAGTTCCACTTCCACACACGGTGACCATCC 1286
Db 1141 TGCTGAAGACAGCCACGCCAGCACCGCGCCCTCCAAGGGAAGTTCCACTTCCACACACGGTGACCATCC 1200

Qy 1287 AGATCAGAGGACTACTCGGAGGACAGCATGAAAGGAGTCAGTGTCAGGCGCCCCCTCAGACT 1346
Db 1201 AGATCAGAGGACTACTCGGAGGACAGCATGAAAGGAGTCAGTGTCAGGCGCCCCCTCAGACT 1260

Qy 1347 GACTGCTCAGCCAGGGCACCA 1366
Db 1261 GACTGCTCAGCCAGGGCACCA 1280

```
            Shiohata,N., Matsumoto,K., Hirano,M., Sano,S., Nomura,R.,
            Yoshikawa,Y., Matsumura,Y., Moriya,S., Chiba,E., Momiyama,H.,
            Onogawa,S., Kaeriyama,S., Satoh,N., Matsunawa,H., Takahashi,E.,
            Kataoka,R., Kuga,N., Kuroda,A., Satoh,I., Kamata,K., Takami,S.,
            Terashima,Y., Watanabe,M., Sugiyama,T., Irie,R., Otsuki,T.,
            Sato,H., Wakamatsu,A., Ishii,S., Yamamoto,J., Isono,Y.,
            Kawai-Hio,Y., Saito,K., Nishikawa,T., Kimura,K., Yamashita,H.,
            Matsuo,K., Nakamura,Y., Sekine,M., Kikuchi,H., Kanda,K.,
            Wagatsuma,M., Murakawa,K., Kanehori,K., Takahashi-Fujii,A.,
            Oshima,A., Sugiyama,A., Kawakami,B., Suzuki,Y., Sugano,S.,
            Nagahari,K., Masuho,Y., Nagai,K. and Isogai,T.
  TITLE     Unpublished
  JOURNAL   Unpublished
            NEDO human cDNA sequencing project
REFERENCE   2  (bases 1 to 1818)
  AUTHORS   Isogai,T. and Yamamoto,J.
  TITLE     Direct Submission
  JOURNAL   Submitted (04-JUL-2002) Takao Isogai, FLJ Project(HRI Team); 2-6-7
            Kazusa-Kamatari, Kisarazu, Chiba 292-0812, Japan
            (E-mail:genomics@hri.co.jp, Tel:81-438-52-3375, Fax:81-438-52-3986)
COMMENT     NEDO human cDNA sequencing project supported by Ministry of
            Economy, Trade and Industry of Japan; cDNA full insert sequencing:
            Research Association for Biotechnology (RAB); cDNA library
            construction: Helix Research Institute (HRI) (supported by Japan
            Key Technology Center etc.); 5'- & 3'-end one pass sequencing: RAB,
            HRI, and Biotechnology Center, National Institute of Technology and
            Evaluation; clone selection for full insert sequencing: HRI and
            RAB; annotation: HRI and RAB.
FEATURES             Location/Qualifiers
     source          1..1818
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="UTERU2009904"
                     /tissue_type="uterus"
                     /clone_lib="UTERU2"
                     /note="cloning vector: pME18SFL3"
BASE COUNT      367 a    516 c    537 g    398 t
ORIGIN

Query Match          50.3%;
Best Local Similarity   99.9%;    Score 813.4;  DB 9;            Length 1818;
Matches  814;  Conservative   0;  Mismatches   1;  Indels   0;  Gaps   0;
                                  Pred. No. 9.4e-164;

Qy    172  TTTCAGGGGCCACTCGCGGCGGGGGGCGGCGGCTGCGGGCGGCGGCTCGGCGGCGGACTGCAT  231
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1  TTTCAGGGGCCACTCGCGGCGGGGGGCGGCGGCTGCGGGCGGCGGCTCGGCGGCGGACTGCAT   60

Qy    232  GGAGGCCAAGGAAGGAAGGAGAAGCAGCATCTGTTGGACGGCCAGGCGGCAATCCGGTCATACACGGG  291
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     61  GGAGGCCAAGGAAGGAAGGAGAAGCAGCATCTGTTGGACGGCCAGGCGGCAATCCGGTCATACACGGG  120

Qy    292  ATCTCTGTGGCAGGAAGGGGCTGGCTGGATTCCTCTCTGCCCTGGACTTGCA  351
           |||||||||||||||||||||||||| ||||||||||||||||||||||||||
Db    121  ATCTCTGTGGCAGGAAGGGGCTGGCTGGAATTCCTCTCTGCCCTGGACTTGCA  180

Qy    352  GGCCCATTGAGCTGGCTGCCCCAGACAACCATCACTGCCATGCCTCAGAAGGGTCCTGACAG  411
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    181  GGCCCATTGAGCTGGCTGCCCCAGACAACCATCACTGCCATGCCTCAGAAGGGTCCTGACAG  240

Qy    412  TCACTGTGACCCCAAGAAGGGGAAGGCCCCAGCCAGCTGTATAGTAGCCCTGTGCCCATCTG  471
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    241  TCACTGTGACCCCAAGAAGGGGAAGGCCCCAGCCAGCTGTATAGTAGCCCTGTGCCCACCTG  300

Qy    472  CCTGTGTGTTCATGATCGGAGAAGTCGTTGGTGGGTACCTGGCACACAGCTTGGCTGTTCAT  531
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    301  CCTGTGTGTTCATGATCGGAGAAGTCGTTGGTGGGTACCTGGCACACAGCTTGGCTGTTCAT  360

Qy    532  GACTGAGGCAGCACACTGCTCACTGACTTTGCCACACATGCTCATCAGCCTCCTTCTCCCT  591
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    361  GACTGAGGCAGCACACTGCTCACTGACTTTGCCACACATGCTCATCAGCCTCCTTCTCCCT  420

Qy    592  CTGGATGTCCTCCCGGCGCCCAGCCACCAAGACCATGAACTTTGGCTGCCAGAGAGCTGAGAT  651
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    421  CTGGATGTCCTCCCGGCGCCCAGCCACCAAGACCATGAACTTTGGCTGCCAGAGAGCTGAGAT  480
```

RESULT 5

BC013478

```
LOCUS       BC013478                2048 bp    mRNA    linear   ROD 07-AUG-2002
DEFINITION  Mus musculus, clone IMAGE:4239648, mRNA.
ACCESSION   BC013478
VERSION     BC013478.1  GI:16975519
KEYWORDS    .
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 2048)
  AUTHORS   Strausberg,R.
  TITLE     Direct Submission
  JOURNAL   Submitted (04-SEP-2001) National Institutes of Health, Mammalian
            Gene Collection (MGC), Cancer Genomics Office, National Cancer
            Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,
            USA
  REMARK    NIH-MGC Project URL: http://mgc.nci.nih.gov
COMMENT     Contact: MGC help desk
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: Jeffrey E. Green, M.D.
            cDNA Library Preparation: Life Technologies, Inc.
            cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
            DNA Sequencing by: Sequencing Group at the Stanford Human Genome
            Center, Stanford University School of Medicine, Stanford, CA 94305
            Web site: http://www-shgc.stanford.edu
            Contact: (Dickson, Mark) mcd@paxil.stanford.edu
            Dickson, M., Schmutz, J., Grimwood, J., Rodriguez, A., and Myers,
            R. M.
            Clone distribution: MGC clone distribution information can be found
            through the I.M.A.G.E. Consortium/LLNL at: http://image.llnl.gov
            Series: IRAK Plate: 25 Row: h Column: 6
            This clone was selected for full length sequencing because it
            passed the following selection criteria: Hexamer frequency ORF
            analysis, Similarity but not identity to protein.
FEATURES             Location/Qualifiers
     source          1..2048
                     /organism="Mus musculus"
                     /db_xref="taxon:10090"
                     /map="FVB/N"
                     /clone="IMAGE:4239648"
                     /tissue_type="Kidney, normal. 5 month old male mouse."
                     /clone_lib="NCI_CGAP_Kid14"
                     /lab_host="DH10B"
                     /note="vector: pCMV-SPORT6"

Qy    652  CTTGGGAGCGCCTCGGTCCTTGCTTGCTTCCTTCTACGTCCAATCTGGGTCGTGGACGGGGGTACTGGGTTGTTGTACCT  711
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    481  CTTGGGAGCGCCTCGGTCCTTGCTTGCTTCCTTCTACGTCCAATCTGGGTCGTGGACGGGGGTACTGGGTTGTTGTACCT  540

Qy    712  GGCGTGTGGGAGCGGGCTGATCTCTGGGGACTATGGAAATTGACGGGGGGGGACCAATGCTGGATCAC  771
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    541  GGCGTGTGGGAGCGGGCTGATCTCTGGGGACTATGGAAATTGACGGGGGGGGACCAATGCTGGATCAC  600

Qy    772  GTCGGGGCTCGGCGTGTGGCCTTGTGAACATCATAATGGGGGTTGGACCCTTCACCAGTCTCTGGGCCA  831
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    601  GTCGGGGCTCGGCGTGTGGCCTTGTGAACATCATAATGGGGGTTGGACCCTTCACCAGTCTCTGGGCCA  660

Qy    832  TGGGCACAGCCCACGGCCACCAACCACCAGGAGGAGAGAACCCCAGCCGTCGGAGCTGCCTT  891
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    661  TGGGCACAGCCCACGGCCACCAACCACCAGGAGGAGAGAACCCCAGCCGTCGGAGCTGCCTT  720

Qy    892  CATCCATGTGATCGGCGACTTTATGCAGAGACCATGGGTTGTTCCTAGTGGCAGCCCTATATATTTT  951
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    721  CATCCATGTGATCGGCGACTTTATGCAGAGACCATGGGTTGTTCCTAGTGGCAGCCCTATATATTTT  780

Qy    952  ATACTTCAAGCCAGAATACAAAGTATGTAGACCCCA  986
           |||||||||||||||||||||||||||||||||||
Db    781  ATACTTCAAGCCAGAATACAAAGTATGTAGACCCCA  815
```

Query Match   46.5%; Score 752.4; DB 10; Length 2048;
Best Local Similarity   79.8%; Pred. No. 1.1e-150;
Matches   947; Conservative   0; Mismatches 211; Indels 28; Gaps 4;

```
BASE COUNT     473 a      557 c      582 g      436 t
ORIGIN

Qy    272 GCAATCCGGTCATACACGGGAGGATCTCTGTGGCAGGAAGGGGCTGGCTGGATTCCTCTCGCC  331
          || |      ||||| ||    |||||||||| |||  || ||| ||| || || |
Db    304 GCCTCCAGGTCCTTCTTAGGATCTCTGTGGAGAGGTGAAGCTTAGCCGGATCCCTC------  358

Qy    332 CGACCTGGCCTGGCACTTGCAGGCCAATTGACCTGGCTGGCTGCCAGAGCAACCATCACTGGCCAT  391
          || ||  |     |   || ||   |   ||  |  ||  ||| ||||||||||||  | | ||||
Db    359 ------CTGTGACCTGCCCGCGGTTGAGCTGGCTGTTCCAGAGCAACCATTATTGCCAT  411

Qy    392 GCTTCAGAAGGGGTCCTGACACAGTCACTGTGACCCCAGAGAAGAGGGGAAGGGCCCAGCCAGCTG  451
          ||  |||||| | ||| || | ||| ||| |||| || |||||| |||| | || || ||   |
Db    412 GCCCAGAAGGATTCGGAAGTCACCCTGACCCCGAGGAAGCAGAGGGCCCGACGCAAACTC  471

Qy    452 TATGTGAGCCCTCTGCCACATCTGCCTGTTGTTCATGATCGGAGAAGTCGTTGGTGGGGTTGGGGTTG  511
          ||||| |||||||||||| ||||||| | |||||| |   |||  || ||| ||||||||||
Db    472 TATGTGGCCTCTGCCATCTGCCTGGTCTCATTGGTCATCTTGTTCATGAATCATTGGTGGGGTTACCTG  531

Qy    512 GCACACAGCTTGGCTGTGTCATGACTGAGCGCAGCACACACCTGCTCACTGACTTTGCCAGCCATG  571
          ||||| || || | |||  |  ||||   ||| | || ||| || ||||||| ||| ||| |||
Db    532 GCACAGGCTTGGCCATCATGACCGACGCTCCGCCGCCTCCACCGGATTTTGCCAGCATG  591

Qy    572 CTCATCAGCCTTCTTCCCCTGGATGTCCCCGGCCCAGCCACCACCAAGACCAATGAAGACTTT  631
          ||||| ||| | ||  | |||  |  |||| |||||||| || || || || |||| |||||||
Db    592 CTCATTAGCCTGTTCTGCCTCTGGGTGTCTTCCCGACCAGCACCAAGACCAATGAACTTC  651

Qy    632 GGCTGGCAGAGACTGAGATCTTGGAGACCCTGGTCTCTGTACTGTCCATCTGGGTGTG  691
          |||||||||||  || || |||||||| ||||  || || || || || ||||||||
Db    652 GGCTGGCCATCGAGATCCTTGGAGCCTTGTTGTCTGTCCTCTCCATCTGGGTTGTG  711

Qy    692 ACGGGGGGTACTGGTGGTTACCTGGCCTCGTGTGGGACGGGCTCGATCTCTGGGGACTATGAAATTGAC  751
          |||||| ||| || ||| || ||||| | |  |||  ||||  |||| || ||||| || ||| | ||
Db    712 ACTGGGGGTCCTGCGGTTACCGTGCAGCGGTTGATTTCTGGAGATTATGAGATCAAA  771

Qy    752 GGGGGGACCATGCATGGGACAACGTCTGGCTGCCGTGCCGTGTCGTGCAAATCATAAATGGGGGTTG  811
          |||||  || |||| ||| | |||||| || |  ||| | ||||| ||||| ||||||||||||||
Db    772 GGGGACCATGCATGGGAACATCATTGGGCTGGCGCTCCACCAGTCTGGTCATGGGGTTG  831

Qy    812 ACCCTTCACCAGTCTGCCATGGGACACAGCGGCACCA---------CCAACCACGAG  862
          || || || || || ||||||||  | ||  |||| ||       | || || ||
Db    832 GCCCTCCATCAGTCTGGACATGGGAACAGCCATGGGAACAGCCATCCTGACGAGCAGCCAG  891

Qy    863 GAGGAGAACCCCAGGGTCCCAGCTGCTCCCTCATCCCATGTGATCGGCGACCTTTATGCAGAGC  922
          |||||||  ||||  | ||| |||||  | |   || |  | || || ||  | ||| || |||
Db    892 CAGCAGAACCCCAGCGTCCGAGCTGCCTTCATCCATGTTGTTGGGGACCTTCTGCAGAGT  951

Qy    923 ATGGGGTGTCCTAGTGGCAGCAGCCTTCCTAATTTTATACTTCAAGCCAGAATACAAGTATGTGAGAC  982
          |||||| || |  | ||| | ||||  ||  ||    | ||  | || || || ||||| |  |||
Db    952 GTGGGGGTACTGGTGGCAGCAGCCTATATCATATTCAAGCCAGAGTACAAGTACGTGGAC  1011

Qy    983 CCCATCTGCACCTTCGTCTTCTCCATCCTGGTCCTGGGGACAACCTTGACCATCCTGAGA  1042
          ||||| |||| || |  | || ||  | |  || |||||  |    ||||||  | ||||
Db   1012 CCCATCTGCACCTTCCTGTTCTCCATCCTGGTCCTGGGGACAACCTTGACCATCCTCAGA  1071

Qy   1043 GATGTGATCCTGGTGTTGATGGAAGGGACCCCCAAAGGGGTGGTTGACTTCACGACTGTTCGT  1102
          ||||| |||||  | |  | || |||||  || |||| || ||| |||||||| || ||  ||
Db   1072 GACGTCATCTTGGTCCTGATGGAAGGGACTCCGAAAGGCGTGGACTTTACCACCGTGAAA  1131

Qy   1103 GATCTGCTTCTGTCGGTGGAGGGGTAGAAGCCCTGCACAGCCTGCATATCTGGGCACTG  1162
          ||| || || | |  | ||| | |  | || ||||||| || ||||| |||| ||| |
Db   1132 AATCTCCTGCTGTCCGTGGATGGGGAAGCCCTCCATAGCCTGCATATCTGGGCCCTG  1191

Qy   1163 ACGGTTGGGCCCAGCCTGTTCTCGTCGTTGTTCACAGAATACAGACGCCGACC  1222
          || ||||||| ||||  | ||||  | |||||||||  | ||  | || || ||
Db   1192 ACAGTGGCCCAGCCGGTGCTGTCTGTCCACATAGCCATATGCTCAGAGCCAG  1251

Qy   1223 GCTGTGCTGAAAGACAGCACGGGATCTGTTGGCAGACGCGCCCTCCAAGGGAAGTTCACACTTCCACACCGTGACC  1282
          | ||| ||||| ||  |  || || || | ||||    |||  | || | ||||| ||| || || |||  |||
Db   1252 GCTGTGCTGAAGGTGGCCCGTGACCGCCTAGGGGAGTTCAATTTCCACACCATGACC  1311
```

RESULT 6

```
LOCUS       RNU50927                1559 bp    mRNA    linear   ROD 20-JUN-1996
DEFINITION  Rattus norvegicus zinc transporter (ZnT-2) mRNA, complete cds.
ACCESSION   U50927
VERSION     U50927.1  GI:1256377
KEYWORDS    .
SOURCE      Norway rat strain=Sprague Dawley.
  ORGANISM  Rattus norvegicus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
            Rattus.
REFERENCE   1  (bases 1 to 1559)
  AUTHORS   Palmiter,R.D., Cole,T.B. and Findley,S.D.
  TITLE     ZnT-2, a mammalian protein that confers resistance to zinc by
            facilitating vesicular sequestration
  JOURNAL   EMBO J. 15 (8), 1784-1791 (1996)
  MEDLINE   96203098
   PUBMED   8617223
REFERENCE   2  (bases 1 to 1559)
  AUTHORS   Palmiter,R.D.
  TITLE     Direct Submission
  JOURNAL   Submitted (07-MAR-1996) Richard D. Palmiter, Howard Hughes Medical
            Institute, University of Washington, Health Sciences Center J661B,
            Box 357370, Seattle, WA 98195-7370, USA
FEATURES             Location/Qualifiers
     source          1..1559
                     /organism="Rattus norvegicus"
                     /strain="Sprague Dawley"
                     /db_xref="taxon:10116"
     gene            1..1559
                     /gene="ZnT-2"
     CDS             52..1131
                     /gene="ZnT-2"
                     /codon_start=1
                     /transl_except=(pos:52..54,aa:Met)
                     /product="zinc transporter ZnT-2"
                     /protein_id="AAB02775.1"
                     /db_xref="GI:1256378"
                     /translation="MARSFFGALWKSEASRIPPVNLPSVELAVQSNHYCHAQKDSGS
                     HPNSEKQRARRKLYVASAICLVPMIGEIGGYLAQSLAIMTDAAHLLTDFASMLISLF
                     SLWVSSRPATKTMNFGWQRAEILGALLSVLSIWVVTGVLVLLAVQRLISGDYEIKDT
                     MITSGCAVAVNIIMGLALHQSGHGHSHSHEDSSQQONPSVRAAFIHVVGDLLQS
                     VGVLVAAYIIYFKPEYKYVDPICTFLFSILVLGTTLTILRDVILVLMEGTPKGVDFTT
                     VKNLLLSVDGVEALHSLHIWALTVAQPVLSVHIAIAQNVDAQAVLKVARDRLQGKFNF
                     HTMTIQIESYSEDMKSCQECQGPSE"
```

```
Db   1252 GCTGTGCTGAAGGTGGCCCGTGACCGCCTAGGGGAGTTCAATTTCCACACCATGACC  1311
Qy   1283 ATCCAGATCGAGAGCTACTCGGAGGACATGAAGAGCTGTCAGGAGTGTCAGGGCCCCTCA  1342
          || ||||| || |||||| ||||  |||| || |||| | || || || ||||| |  | |
Db   1312 ATCCAGATTGAGAGTTACTCTGAGGACATGAAGAGCTGCCAGGAGTGCCAGGGCCCCTCA  1371

Qy   1343 GACTGAGACTGTCAGCCAGGCCACCAAACT-----GGGGCATGAACAGGAGACCTGCAGGTGG--  1395
          || ||| | ||   ||||||||  ||  |||    |||| | ||  | |  |  |  | |||
Db   1372 GAATGACTCCCCAGCCAGACAACACCGTCTCCTCCTCGGAAAGCATGGTCAGGAGACTTGTGGGGTAGAA  1431

Qy   1396 CTGGACTGAGTGTCCCCAGGCCCCAGGCCCAGGACTTTGCCTACCCCA  1441
          || || ||| ||||  ||||| ||| || ||| ||||||| |  ||||
Db   1432 CTGTCCGGTTGGCAGCCGCGGACTCCGGCGAAGGCACAGCCGTGGACCA  1477
```

Query Match   45.4%; Score 734.4; DB 10; Length 1559;
Best Local Similarity   78.5%; Pred. No. 7.6e-147;
Matches   914; Conservative   0; Mismatches 226; Indels 24; Gaps 2;

```
BASE COUNT     353 a      411 c      426 g      369 t
ORIGIN

Qy    218 GCGCGGGACTGCATGAGGAGGCCAAGGAGAAGCAGCACATCTGTTGGCACGCGCACAAATC  277
          |||||||| | |||  |||||||| ||  ||||| ||||| | ||| |||  | ||||
Db      1 GCTCAGGGGCTGCCTCCCTCCGGCCAAGCAATTTTTTCTGCCCTAACCCTGGCCTCC  60

Qy    278 CGGTCATACACGGGGATCTCTGTGGCAGGAGGAAGGGGCTGGGATTCCTCTGCCCGACT  337
```

```
           |||| | || ||| |||
Qy    61 AGGTCTTTCTTGTTGGAGCTCTGTGTGGAAGAGTGAAGCTAGCCGAATCCCGC------  109

Db   338 GGCCTGGACTTGCAGGCCATTGAGCTGGTCTGCCCAGAGCACACCATCACTGCCATGCTCAG  397
Qy   110 -CTGTGAAACCTGCCCTCTGTTGAACTGGCTGTCCAGAGCACCATTACTGTCATGCGCAG  168

Qy   398 AAGGGGTCCTGACAGTCACTGTGACCCCAAGGAAGGGAGGCCCAGCGCGCCAGCTGTATGTA  457
Db   169 AAGGATTCTGGAAGTCACCCCAACTCCGAGAAGCAGAGGGGCCCGACCGCAAACTCTATGTG  228

Db   458 GCCTCTGCCATCTGCCTGTTGTTGTTCATGATCGGAGAAGTCGTTGGGTGGTGGTGGCACAC  517
Qy   229 GCCTCTGCCATCTGCCTCGGTTGTTCATGATATTGGGAGAGATTATTGGGTACCTGGCACAG  288

Qy   518 AGCTTGGGCTTGTCATGACTGACTGAGGCAGCAGCACACCTGCTCACTGACTTTGCCCACATGCTCATC  577
Db   289 AGCTTAGCCCATCATGACCGACGACGGCCGCACCCACCTGCTCACTGATTTCGCCCACATGCTCATT  348

Db   578 AGCCCTCTTTCTCCCCTCTGGGATGTGCTCCCGGGCCAGCCAGACCCAGAACTTTGGGCTGG  637
Qy   349 AGTCTCTTTCTCCCCTCTGGGGTGTGCTTCTCCCGGGGCCAGCCATCGAACTTCGGGTGG  408

Qy   638 CAGAGAGCTGAGAGATCTTGGGAGGACCCCCTGGTGCTCTCGTACTGTCCATCTCGGGTGTCGTGACCGGGG  697
Db   409 CAGCGAGAGCTGAGAGATCCTGAGTCCGTTCATGAGATTATTTGGGTTACCTGGTTGGGTACCGGG  468

Qy   698 GTACTGCTGTGTACCTGCCTGTGTGGAGCGGGCGGCTGATGTATCTCTGGGGACTTTATGCGAGAGCATG  757
Db   469 GTGCTCGTGTGTACCTGGCGTGTACAGCCGGTTGATTTCTGGAGATTATGAGATCAAAGGGGAC  528

Db   758 ACCATGCCTGATCACGTGCGGGCTATATTTTATACTTCAAGCCAGAATACAAGTATGTAGACCCC  817
Qy   529 ACCATGTTGATCACTTCGGGTGGGCTGCCGTCTGTGAACATCATAATTGGGGTTGACCCTTC  588

Qy   818 CACCAGTCTG------GCCATGGGGACACAGCCACGGCACCACCAGCCAGCAGGAG  865
Db   589 CATCAGTGCTGACAATGGACACAGCCATGGGGACCATGAGGACACGAGGACAGGTGGGGG  648

Qy   866 GAGAACACCCACCGTCGGAGCGTCGCCTCATCCATGTGATNCGGGGGACTTTATGCAGGAGCATG  925
Db   649 CAGAACCCAGTGTCCGAGCCTGCCCTTCATTCATGCTGGTTGGGGACCTTCTACAGGAGTGTG  708

Qy   709 GGCGGTCCTCGGGGGCACCCTATATTATTATACTTCAAGCGACTACAAGTACGTGGACCCC  768
Db   926 GGGTGTCCTAGTGCAGCCTATATTTATACTTCAAGCCAGAATACAAGTATGTAGACCCC  985

Qy   986 ATCTGCACCTTCGTCTTCTCCAATCCTGGTCCTGGGGACGACAACCTTGACCATCCTGAGAGAT  1045
Db   769 ATCTGCACCTTCCTCTTCTCCAATCCCTGGTCTTCCAGCGTTAAGCCATCCTGAGAGAT  828

Qy  1046 GTGGATCTCTGGGTGTTGATGGAGGAGCACCCCCAAGGGCCTTGACTTCACGAGCTGTTTCGTTGAT  1105
Db   829 GTGGATCTTGGTTCTCATGGAAGGAGAACTCCCAAAAGGTGTTGGACTACTGTTGAAAAAT  888

Qy  1166 GTGGCCCAGCCTGTCTTCTGTCGTTCCACAGAATACAGACGACCGCCAGGCT  1225
Db   889 CTTCTCCGTCTCCGTGGAGGGGTAGAAGCCCTGCCATCTACAGCCCTGCCATCTCCATATCTGGCGGGCTGACA  948

Qy  1106 CTGCTCCTGCTGTCGGTGGAGGGGGTAGAAGCCCTGCCGCCACTGACG  1165
Db   949 GTGGCCCAGCCCAGCCCGTGCTGCTCGTCTCCACATAGTTTGATCGCCAGGCT  1008

Qy  1226 GTGGCTGAAAGAGAGACCAGCAGCCGCCGCCTCCAAGGGAAGTTCCACTGTTCCACACCGTGACCATC  1285
Db  1009 GTGCTGAAGGTGGCTAGGGAACCGCCTCCAGGGAAGTTCAATTTCCACACACGACCATC  1068

Qy  1286 CAGATCGAGGGACTACTCGGAGGAGGACATGCCAGGGCCCCTCAGAC  1345
Db  1069 CAAAATTGAGAGCTACTCTGAGGACCATGAAGAGAGACTGCCAGGGTGCCTCGGAA  1128

Qy  1346 TGACTGCTGCTCAGCGAGGCACCAACT  1369
```

RESULT 7
```
LOCUS       AK023491
AK023491    2291 bp    mRNA    linear   PRI 01-AUG-2002
DEFINITION  Homo sapiens cDNA FLJ13429 fis, clone PLACE1002500, weakly similar
            to Rattus norvegicus zinc transporter (ZnT-2) mRNA.
ACCESSION   AK023491
VERSION     AK023491.1  GI:10435438
KEYWORDS    oligo capping; fis (full insert sequence).
SOURCE      Homo sapiens placenta cDNA to mRNA, clone_lib:PLACE1
            clone:PLACE1002500.
ORGANISM    Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1
AUTHORS     Isogai,T., Ota,T., Hayashi,K., Sugiyama,T., Otsuki,T., Suzuki,Y.,
            Nishikawa,T., Nagai,K., Sugano,S., Shiratori,A., Sudo,H.,
            Wagatsuma,M., Hosoiri,T., Kaku,Y., Kodaira,H., Kondo,H.,
            Sugawara,M., Takahashi,M., Chiba,Y., Ishida,S., Murakawa,K.,
            Ono,Y., Takiguchi,S., Watanabe,S., Kimura,K., Murakami,K.,
            Ishii,S., Kawai,Y., Saito,K., Yamamoto,J., Wakamatsu,A.,
            Nakamura,Y., Nagahari,K., Masuho,Y., Ninomiya,K. and Iwayanagi,T.
TITLE       Unpublished
JOURNAL     Unpublished
REFERENCE   2 (bases 1 to 2291)
AUTHORS     Isogai,T. and Otsuki,T.
TITLE       Direct Submission
JOURNAL     Submitted (23-AUG-2000) Takao Isogai, Helix Research Institute,
            Genomics Laboratory; 1532-3 Yana, Kisarazu, Chiba 292-0812, Japan
            (E-mail:genomics@hri.co.jp, Tel:81-438-52-3975, Fax:81-438-52-3986)
COMMENT     NEDO human cDNA sequencing project supported by Ministry of
            International Trade and Industry of Japan; cDNA full insert
            sequencing; Research Association for Biotechnology; cDNA library
            construction; 5'- & 3'-end one pass sequencing and clone selection;
            Helix Research Institute (supported by Japan Key Technology Center
            etc.) and Department of Virology, Institute of Medical Science,
            University of Tokyo.
FEATURES             Location/Qualifiers
     source          1..2291
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="PLACE1002500"
                     /tissue_type="placenta"
                     /clone_lib="PLACE1"
                     /note="cloning vector: pME18SFL3"
BASE COUNT      474 a    663 c    647 g    507 t
ORIGIN
```

Query Match 37.2%; Score 601; DB 9; Length 2291;
Best Local Similarity 100.0%; Pred. No. 3.1e-118;
Matches 601; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Db  1129 TAACTCCCTGGCCAGACACCATCT  1152

Qy   966 AATACAAGTATAGAGCCCCATCCTGCCACCTTCGTCTTCTCCAATCCTGGTCCTGGGGACAA  1025
Db     1 AATACAAGTATGTAGACCCCATCTCCACCTTCGTCTTCTCCATCCTGGTCCTGGGGACAA  60

Qy  1026 CCTTGACCATCCTGAGAGATGTGATATCCTGGTTGTTGATGGAGGAGGACCCCCAAGGGCCGTTG  1085
Db    61 CCTTGACCATCCTGAGAGATGTGATATCCTGGTTGTTGATGGAGGAGGACCCCCAAGGGCCGTTG  120

Qy  1086 ACTTCACAGCTGTTCGTTGCTGCTGCTGCTGCGGTGCTGGAGGGGTAGAAGCCCTGCACAGCC  1145
Db   121 ACTTCACAGCTGTTCGTTGATCTGCTGCTGCGGTGCTGGAGGGGTAGAAGCCCTGCACAGCC  180

Qy  1146 TGCATATCTCGGGCACTGACGGTGGCCCCAGCCTGTTCGTGGTTCGTGTTGCGTCCACATCGCCATTGCTC  1205
Db   181 TGCATATTCTGGGCACTGACGGTGGCCCCAGCCTGTTCTGTCTGCCATTGCTC  240

Qy  1206 AGAATACAGACGCGCCCAGGCGTGCTGAAGACAGCAGCCAGCAGCGCAGCGCCTCCAAGGGAAGTTCC  1265
Db   241 AGAATACAGACGCGCCCAGGCGTGCTGAAGACAGCAGCCAGCAGCGCAGCGCCTCCAAGGGAAGTTCC  300
```

RESULT 8

```
Qy   1266 ACTTCCACACCGTGACCATCCAGATCGAGGACTACTCGGAGGACATGAGGACTGTCAGG 1325
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    301 ACTTCCACACCGTGACCATCCAGATCGAGGACTACTCGGAGGACATGAGGACTGTCAGG 360

Qy   1326 CATGCCAGGGCCCCTCAGACTGACTGCTCAGCCACCAGCTCTGGGGCCATGAACAGGAC 1385
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    361 CATGCCAGGGCCCCTCAGACTGACTGCTCAGCCACCAGCTCTGGGGCCATGAACAGGAC 420

Qy   1386 CTGCAGGTGGCTGGACTGAGTGTCCCCAGGACTTTGCCTACCCCAGCTG 1445
          ||||||||||||||||||||||||||||||||||||||||||||||||||
Db    421 CTGCAGGTGGCTGGACTGAGTGTCCCCAGGACTTTGCCTACCCCAGCTG 480

Qy   1446 TGTTATAAACCAGGTGTCCCCTCTCGACCTCTGCCCCACTCCAGGAATGGAGCTCTTCCCA 1505
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    481 TGTTATAAACCAGGTGTCCCCTCTCGACCTCTGCCCCACTCCAGGAATGGAGCTCTTCCCA 540

Qy   1506 GCCTCCCATCTGACTACAGCCCAGGGTGGGGGGGGTATAAAGCTAGTGTGACCCT 1565
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    541 GCCTCCCATCTGACTACAGCCCAGGGTGGGGGGGGTATAAAGCTAGTGTGACCCT 600

Qy   1566 G 1566
          |
Db    601 G 601
```

FEATURES   Location/Qualifiers
     source   1..1440
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="PLACE1002782"
                     /tissue_type="placenta"
                     /clone_lib="PLACE1"
                     /note="cloning vector: pME18SFL3"

BASE COUNT   298 a   420 c   412 g   310 t
ORIGIN

```
Qy   1266 ACTTCCACACCGTGACCATCCAGATCGAGGACTACTCGGAGGACATGAGGACTGTCAGG 1325
Db    301 ACTTCCACACCGTGACCATCCAGATCGAGGACTACTCGGAGGACATGAGGACTGTCAGG 360
```

RESULT 9

```
Query Match     32.2%;  Score 521;  DB 9;  Length 1440;
Best Local Similarity  100.0%;
Matches  521;  Conservative  0;  Mismatches  0;  Indels  0;  Gaps  0;
```

```
Qy   1046 GTGATCCTGGTGTGTTGATGGAAGGGAGGACCCCCAAGGGCGTTGACTTCAGCTGTTCGTGAT 1105
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1 GTGATCCTGGTGTGTTGATGGAAGGGAGGACCCCCAAGGGCGTTGACTTCAGCTGTTCGTGAT 60

Qy   1106 CTGCTCGTGTCTGGGTGGAGGGGGTAGAAGAGCCCCTGCCACAGCCCTGCCATATCCTGGGCACTGACG 1165
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     61 CTGCTCGTGTCTGGGTGGAGGGGGTAGAAGAGCCCCTGCCACAGCCCTGCCATATCCTGGGCACTGACG 120

Qy   1166 GTGGGCCCAGCCCTGTTCTGTCTCGTCTCCACATCGCCATTGCTCAGATATACAGAGACTGGGGCACGGCT 1225
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    121 GTGGGCCCAGCCCTGTTCTGTCTCGTCTCCACATCGCCATTGCTCAGATATACAGAGACTGGGGCACGGCT 180

Qy   1226 GTGCTGAAGACAGCCAGCACAGCACGCGCCGCCTCCAAGGGAAGTTCCAACTTCCACACCGTGACCATC 1285
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    181 GTGCTGAAGACAGCCAGCACAGCACGCGCCGCCTCCAAGGGAAGTTCCAACTTCCACACCGTGACCATC 240

Qy   1286 CAGATCGAGGACTACTCGGAGGACATGAGGACTGTGCAGGGCAATGCCAGGGCCCCTCAGAC 1345
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    241 CAGATCGAGGACTACTCGGAGGACATGAGGACTGTGCAGGGCAATGCCAGGGCCCCTCAGAC 300

Qy   1346 TGACTGCTTCAGCCAGGCCACCAACTGGGGCCATGAACAGGACCTGCAGGTGGCTGGGACTGAG 1405
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    301 TGACTGCTTCAGCCAGGCCACCAACTGGGGCCATGAACAGGACCTGCAGGTGGCTGGGACTGAG 360

Qy   1406 TGTCCCCAGGCCCAGGCAGGACTTTGCCTACCCCAGCTGTGTGTTATAAACCAGGTCCCC 1465
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    361 TGTCCCCAGGCCCAGGCAGGACTTTGCCTACCCCAGCTGTGTGTTATAAACCAGGTCCCC 420

Qy   1466 TCCTGACCTCTCCCCCACTCCAGGAAATGGAGCTCTTCCCAGCCTCCCATCTGACTACAGC 1525
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    421 TCCTGACCTCTCCCCCACTCCAGGAAATGGAGCTCTTCCCAGCCTCCCATCTGACTACAGC 480

Qy   1526 CAGGGTGGGGACTCAGCGGGGTATAAAGCTAGTGTGACCCTG 1566
          ||||||||||||||||||||||||||||||||||||||||||
Db    481 CAGGGTGGGGACTCAGCGGGGTATAAAGCTAGTGTGACCCTG 521
```

FEATURES   Location/Qualifiers
     source   1..11101
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"

BASE COUNT   2403 a   3041 c   3200 g   2448 t   9 others
ORIGIN

```
Query Match     22.5%;  Score 363.6;  DB 6;  Length 11101;
Best Local Similarity  98.9%;  Pred. No. 3e-67;
Matches  366;  Conservative  0;  Mismatches  4;  Indels  0;  Gaps  0;
```

```
Qy   1197 CCATTGCTCAGAATACAGAGACGCCCAGGCTGTGTCGAAGACAGCCAGCAGCCGCCTCCAAG 1256
          ||  |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   8689 CCCCAGCTCAGAATACAGAGACGCCCAGGCTGTGTCGAAGACAGCCAGCAGCCGCCTCCAAG 8748
```

```
Qy    1257 GGAAGTTCCACTTCCACACCGTGACCATCCAGGACTACTCGGAGGACATGAAGG 1316
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    8749 GGAAGTTCCACTTCCACACCGTGACCATCCAGGACTACTCGGAGGACATGAAGG 8808

Qy    1317 ACTGTCAGGCAGATGCCAGGCGCCCCTCAGACTCGACTGCTCAGCCGCCACCAACTGGGGCAT 1376
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    8809 ACTGTCAGGCAGATGCCAGGCGCCCCTCAGACTCGACTGCTCAGCCGCCACCAACTGGGGCAT 8868

Qy    1377 GAACAGGAGACCTGCAGGTGGCTGGGACTGAGTGTCCCCCAGGCCGCAGGCGGACTTTGCCTA 1436
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    8869 GAACAGGAGACCTGCAGGTGGCTGGGACTGAGTGTCCCCCAGGCCGCAGGCGGACTTTGCCTA 8928

Qy    1437 CCCCAGCTGTGTTATAAAACAGGTCCCCCTCCTGACCTCTCCCCCACTCCAGGAANTGGAG 1496
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    8929 CCCCAGCTGTGTTGTTAAACCAGGTCCCCCTCCTGACCTCTCCCCCACTCCAGGAATGGAG 8988

Qy    1497 CTCTTCCCAGCCTCCCATCTGACTACAGCCCAGGTGGGGACTCAGCGCGGGGTATAAAGCTAG 1556
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    8989 CTCTTCCCAGCCTCCCATCTGACTACAGCCCAGGTGGGGACTCAGCGCGGGGTATAAAGCTAG 9048

Qy    1557 TGTGACCCTG 1566
           ||||||||||
Db    9049 TGTGACCCTG 9058
```

```
RESULT 10
LOCUS       AL391650/c
LOCUS       AL391650                 176006 bp    DNA        linear   PRI 28-NOV-2001
DEFINITION  Human DNA sequence from clone RP11-96L14 on chromosome 1, complete
            sequence.
ACCESSION   AL391650
VERSION     AL391650.18  GI:16944045
KEYWORDS    HTG.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 176006)
  AUTHORS   Heath,P.
  TITLE     Direct Submission
  JOURNAL   Submitted (28-NOV-2001) Wellcome Trust Sanger Institute, Hinxton,
            Cambridgeshire, CB10 1SA, UK. E-mail enquiries: clonerequest@sanger.ac.uk
            humquery&sanger.ac.uk Clone requests: clonerequest@sanger.ac.uk
COMMENT     On Nov 15, 2001 this sequence version replaced gi:16605685.
            During sequence assembly data is compared from overlapping clones.
            Where differences are found these are annotated as variations
            together with a note of the overlapping clone name. Note that the
            variation annotation may not be found in the sequence submission
            corresponding to the overlapping clone, as we submit sequences with
            only a small overlap as described above.
            This sequence was finished as follows unless otherwise noted: all
            regions were either double-stranded or sequenced with an alternate
            chemistry or covered by high quality data (i.e., phred quality >=
            30); an attempt was made to resolve all sequencing problems, such
            as compressions and repeats; all regions were covered by at least
            one plasmid subclone or more than one M13 subclone; and the
            assembly was confirmed by restriction digest. The following
            abbreviations are used to associate primary accession numbers given
            in the feature table with their source databases: Em:, EMBL; Sw:,
            SWISSPROT; Tr:, TREMBL; Wp:, WORMPEP; Information on the WORMPEP
            database can be found at
            http://www.sanger.ac.uk/Projects/C_elegans/wormpep This sequence
            was generated from part of bacterial clone contigs of human
            chromosome 1, constructed by the Sanger Centre Chromosome 1 Mapping
            Group. Further information can be found at
            http://www.sanger.ac.uk/HGP/Chr1
            RP11-96L14 is from the library RPCI-11.1 constructed by the group
            of Pieter de Jong. For further details see
            http://www.chori.org/bacpac/home.htm
VECTOR: pBACe3.6
            This sequence is the entire insert of clone RP11-96L14 The true
            left end of clone RP11-231P20 is at 72881 in this sequence. The
            true right end of clone RP11-111D20 is at 72886 in this sequence.
```

```
FEATURES             Location/Qualifiers
     source          1..176006
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="1"
                     /clone="RP11-96L14"
                     /clone_lib="RPCI-11.1"
     misc_feature    169390..169425
                     /note="sequence from uni-directional dGTP big dye
                     terminator reads only."
BASE COUNT      45546 a   43652 c   43353 g   43455 t
ORIGIN
```

```
Query Match          22.5%;  Score 363.6;  DB 9;  Length 176006;
Best Local Similarity  98.9%;  Pred. No. 5.2e-67;
Matches  366;  Conservative  0;  Mismatches  4;  Indels  0;  Gaps  0;

Qy    1197 CCATTGCTCAGAAATACAGACAGCCCCAGGCGCTGTGCTGAAGACAGCCAGCGCGCCTCCAAG 1256
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   24812 CCCCAGCTCAGAAATACAGACAGCCCCAGGCGCTGTGCTGAAGACAGCCAGCGCGCCTCCAAG 24753

Qy    1257 GGAAGTTCCACTTCCACACCGTGACCATCCAGGAGACTACTCGGAGGACATGAAGG 1316
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   24752 GGAAGTTCCACTTCCACACCGTGACCATCCAGGAGACTACTCGGAGGACATGAAGG 24693

Qy    1317 ACTGTCAGGCAGATGCCAGGCGCCCCTCAGACTGACTGAGTGTCTCAGCCGCCACCAACTGGGGCAT 1376
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   24692 ACTGTCAGGCAGATGCCAGGCGCCCCTCAGACTGACTGAGTGTCTCAGCCGCCACCAACTGGGGCAT 24633

Qy    1377 GAACAGGAGACCTGCAGGTGGCTGGGACTGAGTGTCCCCCAGGCCGCCCAGGCGGACTTTGCCTA 1436
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   24632 GAACAGGAGACCTGCAGGTGGCTGGGACTGAGTGTCCCCCAGGCCGCCCAGGCGGACTTTGCCTA 24573

Qy    1437 CCCCAGCTGTGTTATAAAACAGGTCCCCCTCCTGACCTCTCCCCACTCCAGGAATGGAG 1496
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   24572 CCCCAGCTGTGTTGTTGGTTAAAACAGGTCCCCCTCCTGACCTCTCCCCACTCCAGGAATGGAG 24513

Qy    1557 TGTGACCCTG 1566
           ||||||||||
Db   24452 TGTGACCCTG 24443.
```

```
RESULT 11
LOCUS       MMU76007
LOCUS       MMU76007                 1952 bp    mRNA        linear   ROD 02-JAN-1997
DEFINITION  Mus musculus zinc transporter ZnT-3 (ZnT-3) mRNA, complete cds.
ACCESSION   U76007
VERSION     U76007.1  GI:1763373
KEYWORDS    .
SOURCE      Mus musculus.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 1952)
  AUTHORS   Palmiter,R.D., Cole,T.B., Quaife,C.J. and Findley,S.D.
  TITLE     ZnT-3, a putative transporter of zinc into synaptic vesicles
  JOURNAL   Proc. Natl. Acad. Sci. U.S.A. (1996) In press
REFERENCE   2  (bases 1 to 1952)
  AUTHORS   Palmiter,R.D., Cole,T.B., Quaife,C.J. and Findley,S.D.
  TITLE     Direct Submission
  JOURNAL   Submitted (23-OCT-1996) Howard Hughes Medical Institute, University
            of Washington, Box 357370, Seattle, WA 98195, USA
FEATURES             Location/Qualifiers
     source          1..1952
                     /organism="Mus musculus"
                     /db_xref="taxon:10090"
                     /chromosome="5"
     gene            1..1952
                     /gene="ZnT-3"
```

```
CDS             129..1295
                /gene="ZnT-3"
                /function="zinc transporter"
                /codon_start=1
                /product="ZnT-3"
                /protein_id="AAB39731.1"
                /db_xref="GI:1763374"
                /translation="MEPSLATGGSETTRLVSARDRSSAGGGLRLKSLFTEPSEPLPEE
                PKLEEMAFHHCHKDPVPQSGLSPERVQARRQLYAACAVCFIFMAGEVVGGYLAHSLAI
                MTDAAHLLADIGSMLASLFSLWLSTRPATRNTMTFGWHRSETLGALASVVSLWIVTGIL
                LYIAFLRLLHSDYHIEAGAMLITASIAVCANLLMAFVLHQGTGAPHSHGSTGAEYAPLE
                EGHGYPMSLGNTSVRAAFVHVLGDLLQSFGVLAASILIIYFKPQYKVADPISTFLFSIC
                ALGSTAPTLRDVLLVLMEGAPRSVEFEPVRDTLLSVPGVRATHDLHIWALTLTHVAS
                AHLAIDSTADPEAVLAEASSRLYSRFGFSSCTLQVEYQPEMAQCLRCQEPSQA"
BASE COUNT      326 a   615 c   542 g   469 t
ORIGIN
```

```
Query Match              21.7%;  Score 350.6;  DB 10;  Length 1952;
Best Local Similarity    60.5%;  Pred. No. 1.3e-64;
Matches   618;  Conservative    0;  Mismatches 364;  Indels  39;  Gaps   1;

Qy   360 AGCTGCTGCCAGAGCCAACCATCACTGCACTCTCAGAAGGTCCTTGACAGTCACTGTG 419
         |||| ||         ||   ||    ||    ||       |||    |||   |||
Db   265 AGCTTGAGGGGATGGGCCTTCCACCACTGCCTGCCACCACTGCCTCATTGCCCTCT 324

Qy   420 ACCCCAAGAGAAGGGGAAGCCCCAGCGCCGCCAGCGTGTATGTAGCCTCTGCCATCTGCCTCGTTGT 479
         ||                ||   ||    ||   |||    ||   |||    |||   |||   |||
Db   325 CTCCAGAGAGGGTGCAAAGCCCGGAGGCAGCTGTATCCGCGCTTGTGTGTTCATCT 384

Qy   480 TCATGATCGGGAGAAGTCGTTGGTGTGGGTACCTGGCACACAGCTTGGCTGTCATGACTGACG 539
         ||   ||  |||       |||  |||||        ||    ||  |||   |||   ||
Db   385 TCATGGCCGGGGAGGTGGTGGGTGGTTATTTAGCACAGCGTTGGCCATCATGACTGATG 444

Qy   540 CAGCACACTGCTCACTTTGCCAGCATCAGCCTCTTCTCCCCTCTGGATGT 599
         | ||||||||||   |   |||    ||       ||       ||
Db   445 CCGCCACACTGCTAGCAGACAATAGGCGACATGTGGGCAGCTGTTTTCCCTCTGGCTTT 504

Qy   600 CCTCCCGGCCCAGCCACCAAGACCATGAACTTTGGCGTGCGACAGAGCTGAGAGATCTTGGGAG 659
         |  | |||||||||||||  ||||||||||  ||||  |||   ||||  ||   |||||||
Db   505 CCACTCGGCCGCAGCCACCCGACCCACCATGAACCTTTGGCTGTGGCGACTGCGACGGTCGGAG 564

Qy   660 CCCTGGTCTCTGTACTGTCTGGGTCGTCGTGTGTGACGGGTACTGCGTGTTACTGGCTGTGTGG 719
         |||||   |  |||      |||   ||||  |||   |||   ||||    ||  |||||||
Db   565 CTTTTGCCTGTGTGGTCTCCCTCTTGATAGTCACTGCCATCCTCCTGTTACTGGCCTTCC 624

Qy   720 AGCGGCCTGGATCTCTGGGGACTATGAAAATTGACGGGGGACCATGTCGAATCACGTCGGGGCT 779
         |  |||||||      |||     |||      ||||   ||||||    ||     ||  |||
Db   625 TCCGCCTGCTTCATAGTGACTATCACATTGAGGGGGGTGCCATGCTGATCACAGCCAGCA 684

Qy   780 GCGGCTGTGGCTGTGAACATCATAATGGGGTTGACCCTTCACCAGTCTGTGGCCATGGGCACA 839
         ||    ||| |   |  |||    |   |||||||||   |   |||   ||  |||||   ||
Db   685 TCGCTGTCTGTGCCAATCTGCTTAATGGCCTTTGTGCTGCACCAGACTGGAGCCCCACA 744

Qy   840 GCCACGGGCGCACCAACCCAGCAGGA-------- 864
         |||||| |  |  |||     |||||
Db   745 GCCATGGATCTACAGTGCACAGGATGCACCCTCTAGAGGAAGGGGATCAGCGCAAATGT 804

Qy   865 ----GGAGAACCCACCAGCGTCCGAAGGTCGCCTTCTCCATCCGTGGACACTTTATGCCAGA 920
             || |  |  || |||||||   ||      ||    ||||        |||||   ||||
Db   805 CCCTGGGAAACACGAGTGTCCGGGGCGTGCCTTTGTTCATGTTCTGGGAGACCTTCCTCAGA 864

Qy   921 GCATGGGGTGTCCTAGTGGCAGCCTATATTTTTATACTTCAACGCCAGGAATACAAAGTAGTATGTAG 980
         || |||    |||  ||  | ||||     ||      ||   ||||    ||  ||||    |||
Db   865 GCTTTGGGGTCCTGCGAGCCGATCCTACTTCAAGCCTACAAGGTAGCTG 924

Qy   981 ACCCCAATCTGCACCTTCGCTTCTTCCAATCGTGGGCTCTGGGACAACTTGACCATCTCTGA 1040
         ||||||||    |  ||    |    |||     ||    ||||    ||   |||  |||
Db   925 ACCCCATCAGCACCTTCCTCTTCTCCATCTGTGCCCTTGGGTCTACCGCAACCCTGCGAGATGT 984

Qy  1041 GAGATGTGATCCTGGTGTTGATGGAAGGAGGGCCCCGAAGTGTTGAGTTCGAGCCCGTGCGTGTTC 1100
         ||||| | |||||   |||   |||||||   |||   | |  | |  |||  ||| ||||
Db   985 GAGAGTGTCTCCGTATTCTGATGGAAGGGACCCCTCGGAATGTGGGAATTGAACCCGTAA 1044

Qy  1101 GTGATCTGCCTGCTGTCGGTGGAGGGGTAGAGAGCCCTGCCACAGCCTGCATATCTGGGCAC 1160
         |||||||||    ||||  ||   || |||  |||  ||||||  ||   ||   ||||| |
Db  1045 GGGACACCTTGTTGTCCGTGCCGTGCCCGGGAGTCCGAGCAACCCATGAGATCCCATCTGTGGGGCCC 1104

Qy  1161 TGACGGTGGCCCAGCCTGTTCTGTCTGGTGCCACATCGCCATTGCTCAGAATACAGACGCC 1220
         ||||| || |||||||||   |||| | ||||||   |  |||| ||| ||||  |||||
Db  1105 TGACGCTTACTTACCACGTTGCCTCTGCCACACCTGGCTATTGACTCCACGGTGACCCTG 1164

Qy  1221 AGGCTGTGCTGAAGACAGCAGCCAGCGACCGCCGCCTCCAAGGGAAGTTCCACTTCCACACCGTGA 1280
         |||||||   ||||    ||  ||  |    |||| ||  || |  |||| | |||  |||
Db  1165 AAGCTGTCCTGGGTGAAGCCTCATCCCGGTGTTGGATTCTCAGCTGCA 1224

Qy  1281 CCATCCAGATCGAGGACTACTCGGAGACACATGAAGAGACTCTCCGAGA 1340
         ||||||||||||||||||| |||||||| ||||| ||||||||||||||
Db  1225 CCCTGCCAGGTGGAGCAATACCAGCCTGAGATGGCCCAGTGTCTGCGCTGCCAGGAGCCCT 1284

Qy  1341 C 1341
         |
Db  1285 C 1285
```

```
RESULT 12
LOCUS       HSU76010
HSU76010                2000 bp    mRNA    linear   PRI 02-JAN-1997
DEFINITION  Human putative zinc transporter znT-3 (znT-3) mRNA, complete cds.
ACCESSION   U76010
VERSION     U76010.1  GI:1763375
KEYWORDS    .
SOURCE      Homo sapiens.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 2000)
  AUTHORS   Palmiter,R.D., Cole,T.B., Quaife,C.J. and Findley,S.D.
  TITLE     ZnT-3, a putative transporter of zinc into synaptic vesicles
  JOURNAL   Proc. Natl. Acad. Sci. U.S.A. (1996) In press
REFERENCE   2  (bases 1 to 2000)
  AUTHORS   Palmiter,R.D., Cole,T.B., Quaife,C.J. and Findley,S.D.
  TITLE     Direct Submission
  JOURNAL   Submitted (23-OCT-1996) Howard Hughes Medical Institute, University
            of Washington, Box 357370, Seattle, WA 98195, USA
FEATURES             Location/Qualifiers
     source          1..2000
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
     gene            1..2000
                     /gene="ZnT-3"
     CDS             84..1250
                     /gene="ZnT-3"
                     /function="zinc transporter"
                     /codon_start=1
                     /product="ZnT-3"
                     /protein_id="AAB39732.1"
                     /db_xref="GI:1763376"
                     /translation="MEPSPAAGGLETTRLVSPRDRGGAGGSLRLKSLFTEPSEPLPEE
                     SKPVEMPFHHCHRDPLPPGLTPERLHARRQLYAACAVCFVFMAGEVVGGYLAHSLAI
                     MTDAAHLLADVGSMMGSLFSLWLSTRPATRNTMTFGWHRSETLGALASVVSIWMVTGIL
                     LYIAFVRLLHSDYHIEGGAMLITASIAVCANLLMAFVLHQGPPHSHGSRGAEYAPLE
                     EGPEQPLPLGNTSVRAAFVHVLGDLLQSFGVLAASILIIYFKPQYKAADPISTFLFSIC
                     ALGSTAPTLRDVLRILMEGTPRNVGFEPVRDTLLSVPGVRATHELHIWALTLTYHVAS
                     AHLAIDSTADPEAVLAEASSRIYSRFGFSSCTLQVEYQPEMAQCLRCQEPPQA"
BASE COUNT      301 a   659 c   590 g   450 t
ORIGIN
```

```
Query Match              21.4%;  Score 346;  DB 9;  Length 2000;
Best Local Similarity    60.7%;  Pred. No. 1.3e-63;
Matches   608;  Conservative    0;  Mismatches 355;  Indels  39;  Gaps   1;

Qy   379 CCATCACTGCACTGCCATCAGCTCAGAAGGTCCTGACAGTCACTGTGACCCCAAGAGAAGGGGAAGC 438
         ||    |||  |  |     ||   ||||||         |||   ||||||||   |  |||  |||
Db   239 CCACCACTGCCACAGGGACCCCCGCCTTCCCGCCGCCTTACCCTGAGAGGCTGCATGC 298

Qy   439 CCAGCGCCAGCGTGTATGTAGCCTCTGCCATCTGCCTCGTTGTTCATGATCGGGAGAAGTCGT 498
```

```
ORGANISM  Homo sapiens
          Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
          Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 2101)
AUTHORS   Strausberg,R.
TITLE     Direct Submission
JOURNAL   Submitted (23-APR-2002) National Institutes of Health, Mammalian
          Gene Collection (MGC), Cancer Genomics Office, National Cancer
          Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,
          USA
REMARK    NIH-MGC Project URL: http://mgc.nci.nih.gov
COMMENT   Contact: MGC help desk
          Email: cgaps-r@mail.nih.gov
          Tissue Procurement: Miklos Palkovits, M.D., Ph.D.
          cDNA Library Preparation: Michael J. Brownstein (NHGRI) & Shiraki
          Toshiyuki and Piero Carninci (RIKEN)
          cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
          DNA Sequencing by: Institute for Systems Biology
          http://www.systemsbiology.org
          contact: amadan@systemsbiology.org
          Anup Madan, Jessica Fahey, Erin Helton, Mark Ketteman, Anuradha
          Madan, Stephanie Rodrigues, Amy Sanchez and Michelle Whiting

          Clone distribution: MGC clone distribution information can be found
          through the I.M.A.G.E. Consortium/LLNL at: http://image.llnl.gov
          Series: IRAK Plate: 34 Row: m Column: 1
          This clone was selected for full length sequencing because it
          passed the following selection criteria: matched mRNA gi: 4508042.

FEATURES             Location/Qualifiers
     source          1..2101
                     /organism="Homo sapiens"
                     /db_xref="LocusID:7781"
                     /db_xref="taxon:9606"
                     /clone="MGC:26355 IMAGE:4837595"
                     /clone_lib="NIH_MGC_97"
                     /lab_host="DH10B"
                     /note="Vector: pBluescript"
                     /tissue_type="testis"
     CDS             180..1346
                     /codon_start=1
                     /product="solute carrier family 30 (zinc transporter),
                     member 3"
                     /protein_id="AAH28358.1"
                     /db_xref="GI:20306868"
                     /translation="MEPSPAAGGLETTTRLVSPRDRGGAGGSLRLKSLFTEPSEPLPEE
                     SKPVEMPHHICHHRDPLPPGLTPERLHARRQLYAACAVCFVFMAGEVVGGYLAHSLAI
                     MTDAAHLLADVGSMWGSLFSIWLSTRPATRTMFFGWHRSETLGALASVVSLWMVTGIL
                     LYIAFVRLLHSDVHIEGGAMLLTASIAVCANLLMAFVLHQAGPPHSHGSRGAEYAPLE
                     EGPEEPLPIGNTSVRAAAVHVLGDLLLQSFGVLAASILIYFKPQYKAADPISTFLFSIC
                     ALGSTAPILRDVLRILMEGTPRNVGFEEPVRDTLLLSVPGVRATHELHLWALTLTYHVAS
                     AHLAIDSTADPEAVLAEASSRLYSRFGFSSSCTLQVEQYQPEMAQCLRCQEPPQA"
BASE COUNT      315 a    692 c    627 g    467 t
ORIGIN
```

RESULT 13

```
LOCUS       BC028358                2101 bp    mRNA    linear   PRI 25-APR-2002
DEFINITION  Homo sapiens, solute carrier family 30 (zinc transporter), member
            3, clone MGC:26355 IMAGE:4837595, mRNA, complete cds.
ACCESSION   BC028358
VERSION     BC028358.1  GI:20306867
KEYWORDS    MGC.
SOURCE      Homo sapiens.
```

```
Query Match             21.4%;  Score 346;  DB 9;  Length 2101;
Best Local Similarity   60.7%;
Matches   608;  Conservative    0;  Mismatches  355;  Indels   39;  Gaps   1;
```

```
Qy    379 CCATCACTGCCATCCTCAGAAGGGTCCTGACAGTCACTGTGACCCCAAGAGAAGGGGAAGGC  438
          ||| ||||||||   || | ||||| ||| | |     | | || ||||   ||  |||||
Db    335 CCACCACTGCCACAGGGACCCCCCTTCCCGCGCGGCCTTACCCCTGAGAGGCTGCATGC  394

Qy    439 CCAGGCAGCTGTATGTAGCCCTCGCCATCTGCCTGTGTTCATGATCCGGAGAGTCGT  498
          ||  ||  ||   |  |||    | | |||  || |||   |||||| || |  ||||
Db    395 ACGGAGGCAGCTATATGCTGCCCTGTGCCGTTGGTTGCTTGTTCATGGCTGGGAGGTGGT  454

Qy    499 TGGTGGGTACCTGGCACACACGCTTGGCTGTGCATGAGCAGCACACACCTGCTCACTGA  558
          ||||||| |||||||||||||||||||||||| || |||||||||||||||| || |||
Db    455 CGGCGGGTATCTGGCCCATCACTGGGCATCATGACGATGAACGCCAGCCCACTTGCTGGGGA  514

Qy    559 CTTTGCCAGCATGTCATCAGCCCTCTTCTCCCTCTGGATGGTCCCGGGCCCAGCCACCAA  618
          || |||||||||| ||| ||||| || |||| | || | ||| |  ||||  ||| ||||
Db    515 TGTGGGCAGCATGATGGGCAGCCTCTTCTCCATCTGGCTCTCCACCCGTCCAGCCACCG  574
```

Alignment blocks (Qy = query, Db = BC028358):

```
Db    299 ACGGAGGCAGCTATATCGTGCTTGCCTGTGCCGTTGGTTGCTTGTTGTCTTCATGGCTGGGAGGTGGT  358
Qy    499 TGGGTGGGTACCTGGCACACACGCTTGGCTGTGCATGAGCAGCACACACCTGCTCACTGA  558

Db    359 CGGCGGGTATCTGGCCCATCACTGGGCATCATGACGATGAACGCCAGCCCACTTGCTGGGGA  418
Qy    559 CTTTGCCAGCATGTCATCAGCCCTCTTCTCCCTCTGGATGGTCCCGGGCCCAGCCACCAA  618

Db    419 TGTGGGCAGCATGATGGGCAGCCTCTTCTCCCTCTGGCTCTCCACCCGTCCAGCCACCG  478
Qy    619 GACCATGAACTTTGGCTGGCACAGAGAGCTGAGAATCTTGGGAGCCCTGGTCTCTGTACTGTC  678

Db    479 CACCATGAACCTTTGGCTGCAGGGGGGTGCCACGGGGGTGCCATGGGTCTTGGGCCTTCCTC  538
Qy    679 CATCTGGTTCGTGACCGGGGGTACTGGGTGGTTCAGAGACTCTGGTGTGTCTGTGTGGGTC  738

Db    539 CCTCTCGGATGGTCACTGGCAATCCTCCTGTACCTGGGCCTTCGTCGCTGCCTGCTCACCAGCGA  598
Qy    739 CTATGAAATTGACGGGGGACCATGGCTGCTGACTCACGTCGGGCTCGTGGCTGTGTGAACAT  798

Db    599 CTACCACATCGAGGGGGGTGCCATGCTGCTGACCGCCAGCATCGCAGTCTGTGCCAACCT  658
Qy    799 CATAATGGGGTTGACCCTTCACCAGTGTGGCACATGGGCACACGGGCACCA--------  851

Db    659 GTTAATGGCCTTTGTGCTGCACCAGGCTGGTCACCAGGGCCGGGCCCCCACGGGGGAGC  718
Qy    852 ----------CCAACCAGGAGGAGGAGAGAACCCCAGCGT  879

Db    719 AGAGTATGCACCGCTGGAGGAGGGGCCTGAACAGCCCCTGCCCCTGGGGAACACCAGCGT  778
Qy    880 CCGAGCTGCCTTCATCCATGTGATCGGCGACTTTATGCAGAGCCATGGGGTGTCCTAGTGGC  939

Db    779 CCGGGCGGCCATTGTGCACGTGCTGGGGGACCTCCTGCTCCAGAGCTTTGGGGTACTGGCTGC  838
Qy    940 AGCCTATATTTTATACTTCAAGCCACAGATAACAAGTATGTAGAACCCATCTCACCTTCGT  999

Db    839 CTCCAACCTCATCTACTTCAAGCCTCAAATACAAAGGCAGCCGACCCCATCAGCCAACCTTCCT  898
Qy   1000 CTTCTCCATCTGTGCTCTGGGGGACAACACCATCCTTGACCATGGATCTGGTTGTTGTT  1059

Db    899 CTTCTCCATCTGTGCCCTTGGAATCCACGCGCTTCCACCTGCGGTTCGAGCCTTTCGA  958
Qy   1060 GATGGAAGGGAGGGACCCCAAGGGCGTTGACTTCACAGCTGTTCGGTGATCTGCGTCGTCGGT  1119

Db    959 CATGGAAAGGTACCCCGCCGCAAATGTGGGTTCGGGAATAACGTGTGTTGTCGGT  1018
Qy   1120 GGAGGGGGTAGAAGCCCTGCACAGCCTGCCCATATCTGGGCCACTGACGGTGCCCCAGCCTGT  1179

Db   1019 GCCAGGAGTCCGGGCAACCCATGAGCTGACCCTGTGGGCCCTTACGCTGTCCACTTACCATGT  1078
Qy   1180 TCTGTCTGTCCACATCCTGGTCTGGGGACAACACCATCCTGAGAATACAGCAGCGGGCTCGTGTCTGTGAAGAGCAGC  1239

Db   1079 TGCCTCTTGCACACCTGGCCCATGGCCCATCCCACCCCGTCCCCTGAAGCCGTCCTGCTGAAGC  1138
Qy   1240 CAGCAGCGGCCTCCAAGGGAAGTTCCACTCCACACGTGGACATCCAGATCGAGGACTA  1299

Db   1139 CTCCATCCCGGCTTCTACTCCCGGTTTGGATTCTCACGTCGACCTTGCAGGTCGAGCAGTA  1198
Qy   1300 CTCGGAGGACAATGAAGGACTGTCAGGGACCATGCCCAGGGGCCCCTC  1341

Db   1199 TCAGCCGGAGATGGCCCAGTGCCTGCGCTGCCAGGAACCCCC  1240
```

```
RESULT 14

LOCUS       AX061210                2823 bp    DNA     linear   PAT 22-JAN-2001
DEFINITION  Sequence 57 from Patent WO0078953.
ACCESSION   AX061210
VERSION     AX061210.1  GI:12406346
KEYWORDS    .
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 2823)
  AUTHORS   Lal,P., Yang,J., Yue,H., Hillman,J.L., Tang,Y.T., Bandman,O.,
            Burford,N., Baughn,M.R., Azimzai,Y., Lu,D.A., Au-Young,J. and
            Patterson,C.
  TITLE     Human transport proteins
  JOURNAL   Patent: WO 0078953-A 57 28-DEC-2000;
            Incyte Genomics, Inc. (US)
FEATURES             Location/Qualifiers
     source          1..2823
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
BASE COUNT      838 a    599 c    621 g    765 t
ORIGIN

Query Match             19.1%;  Score 309.6;  DB 6;  Length 2823;
  Best Local Similarity   57.9%;  Pred. No. 8.4e-56;
  Matches  574;  Conservative    0;  Mismatches  409;  Indels    9;  Gaps    1;
```

```
Qy     619 GACCATGAACTTTGGCTGGCAGAGAGCTGAGATCTTGGGAGGCCCGGGTCTCTGTACTGTC 678
           ||||||  ||||||||| || |||||||| || ||||| || |||| ||| ||| || ||
Db     575 CACCATGACCTTTGGCTGCTGGCAGAGACTCTGGGGGCTTTGGCCTCTGCTGTGTCTC 634

Qy     679 CATCTGGGTCGTGACGGGGGGGTACTGGTGTACCTGGCGGCTTGGAGGGCCGGTTCTTGTC 738
           ||||| || |||  ||||||  || || ||||  ||||||  ||||  || || ||  ||
Db     635 CCTCTGGATGGCACTGGTGTACCTGTACCTGGCCTTGGGGGCTTTGGCCTGTGCTGCACAGCGA 634

Qy     739 CTATGAAATTGACGGGGGGAGGACCATTGATCACGTCGCGCTTGTGGCTGTGAACAT 798
           |||||||| || || |||| || ||||| ||||||| || || || |||||| || |
Db     695 CTACCACATCGAAGGGGGGGTTGCCATGCTCCTGTGTGACCTTGGACTCTGCCTGTGCCAACCT 754

Qy     799 CATAATGGGGTTGACCCTTCACCAGTCTCTGGCGTCGCGCGCTGTGTGGCGTGTGAA 851
           ||||  || || || |||||||||| |||| || |  |||  || || |||| |||
Db     755 GTTAATGGGCCTTTGTTGCTCACCAGGCTGGGGGCCGGTTCAGCGGTACTGGGGAGC 814

Qy     852 ------------------------CCAACCAGGAGGAGGAGAAGAACCCCAGCGT 879
                                   ||||||||| || |||| || || ||||| ||
Db     755 AGCCTATATTTTTATTCAAGCCAGAATACAAGTATGTAGACCCCATACTCACCTTCGT 814

Qy     880 AGAGTATGCACGGCTGGAGGAGGGGCCTGAAGAGCCCTTGGTTGTCCTAGTGGC 939
           |||| ||| || || ||||| || |||||| || ||||  || || || || ||
Db     815 CCGAGCTGCCTTCATCCATGATGCGGACTTTATGCAGAGGCATGGGGTGTCCCAGTGGC 874

Qy     940 AGCCTATATTTTTATTCAAGCCAGAATACAAGTATGTAGACCCCATACTCACCTTCGT 999
           ||||  || ||  || || ||||||  || |||| || || || ||||| |||| |
Db     875 CCCGGGCGGCCATTGTGCACGTGCGTGGGGGGACCTCCTGCACTTTGGGGTACTGGCTGC 934

Qy    1000 CTTCTCCATCCTGTTCCTCTGGGACACAACCTTGACCATCCTGAGAGAATGTGAATCCTGGTTGTT 1059
           ||| || ||||||| ||| |||| || || || || || || |||||| || || || || ||
Db     935 CTCCAATCCTCATCTACTTCAAGCCAGAATACAAGGCAGCACCCCATCAGCACACCTTTCCT 994

Qy    1060 GATGGAAGGGACCCCCAAGGGCGTTGACTTCACAGCTGTTCGTGATCCTGTTGTCGGT 1119
           ||||||||  || ||||||||| || || || |||| || || ||| || || || ||
Db     995 CTTCTCCATCTGTGCCCTTGGCCCTTCCACCGTCCCAGCACGTTCTTCGAATCCT 1054

Qy    1120 GGAGGGGGTAGAAGCCCTGCACAGCCTGCATATATCTGGGACACTGACGGTGCCCAGCCTGT 1179
           ||||||  || ||||| || |||| ||  ||||||||| || |||||||  ||||||||||
Db    1055 CATGGAAGGTACCCCCCGCCAAGTGTGGGGTTCGAACCTGTGGGGATAACCTGTTGTGCGGT 1114

Qy    1180 TCTGTCTGTCTCCACATCGCCATTGCTCAGAATACACGACGCCCCAGCGTCTGTGTCTGAAGACAGC 1239
           || || || || || || || ||||| || || || |||| || |||||| || || || || |||
Db    1115 GCCAGAGTCCCGGGCGGCCAAACCCAATGAGTCTGTCGGGGGCCCTTACGGCTCACTTACCATGT 1174

Qy    1240 CAGCAGCGCCTCCCAAGGGAGAAGTTCCACTTCCACACGTGACCCATCCAGATCGAGGACTA 1299
           |||||||  ||| ||||||| || || || || || || || ||| || |||| ||| ||
Db    1175 TGCCTCTTGCACACTGGGCCATCGACTCCACGCTGACCCTGAAGCCGTCCTGGCTGAAGC 1234

Qy    1300 CTCGGAGGACATGAAGAGGACTTCTCAGCGCATGCCAGGGCCCCTC 1341
           || ||||| ||||| || || ||  |||| |  ||||||| ||||
Db    1235 CTCATCACCCGGCTCTACTCCCGGTTTGGAATTCTCCAGCTGCACCCGCACCTGCAGCAGTA 1294

Qy    1295 TCAGCGAGGAGATGGGCCCAGTGCCTCGCGCTGCCAGGGAACCCCC 1336
           ||||| || || || ||||||||| ||||||||||||| |||||
Db    1295 AAGCCCAGGATAAAAATAGCCGAGACCCAAATCTGCACATTCCATCCTGGTCTG 1336
```

```
Qy     359 GAGCTGGCTGCCCAGAGCAACCATCACTGCCATGCTCAGAAGGTCCTGACAGTCACTGT 418
           ||||||  ||| |||||||| || || ||||||||||| || || || |||| || ||
Db     439 GAGCTGGACTCAGGAGGCATGTACCACTGCCACAGTGCCCAGAAAAAGGG 498

Qy     419 GACCCCAAGAGAAGGGGAAGGGCCCAGCCGTGTCTATGTGAGCTCTGCCATCTGCCTTG 478
           || || ||||| || || |||| || || || |||||| || || || |||| || |
Db     499 GCGAAATGAGTACGGCCTATGCCAAGTGGAAACTCTGTTCTGCTTCAGCAATATGCTTCATT 558

Qy     479 TTCATGATCGGAGAAGTCGTTGGTTGGGGGTCACACAGCTTGGCTGGTTGTCATGACTGAC 538
           |||||||||||||| || ||  || ||||| ||||||| || || ||| || || || ||
Db     559 TTCATGATTGCAGAGGTCGTGGGGGCACACAATTGCTGGAGTCGTGGGGTCTAGTGGTACCTGGT 618

Qy     539 GCAGCACACCTGCTCACTGACTTTGCCACCATCAGCAGCCTCATCAGCGTCTCTTCCCTGGATG 598
           ||| ||| || || |||||||||||||| || |||||| ||||||| || || || || ||
Db     619 GCTGCCCACCTTGTTAATGACCTGACCAGTTTCCTGCTCAGTTCTTCTCCCCTGGGTTG 678

Qy     599 TCCTCCCGGCCCAGCCACCAACCATCAAACTTTGGGGTTGGGCAGAGAGTTGAGAGTTCTGGGA 658
           || || ||| || || || ||||| ||| ||||||| |||| || ||| || || || ||||
Db     679 TCATCGAAGCCTCCTCCTTCAAGCGCCTGTTCTGCTTCAGCAATATGTTCAT 738

Qy     659 GCCCTGGTTCTCGTTGTACTTCCATCTGGGTCGGTGTGACCGGGGTGTACTGGTGTACCTGTGTG 718
           |||||  ||| || || || ||||||  ||| ||| ||||||| || ||||| |||| || ||
Db     739 GCCCTGCTGTCTCATCCACCTGTGTGCANCTGCTGTTCTGTGTTGACTGGTGTACCCTGGCATGT 798

Qy     719 GAGCGGCTGCGATCTCTGGGGGACTATGAAAATGACGGGGGGAGGGCATGTCAGCAGGTCAT 778
           |||| ||||| || ||| ||||| || |||||||||||||| || ||| || || ||||
Db     799 GAGGCCTGCTGCTGTGATCCTGATTACCAGAGTCAGGGACGACTGTGATCATCGTTTCAGC 858

Qy     779 TCGCGCTGTGGCGTGTGGACATCATATAAATGGGTTGGGTCGGGCTGCCATGGCAC 838
           || || || ||  || || |||||||| ||||||||||| || ||| ||| ||||
Db     859 TGGCGACTGGGCGGCGCCAACATTGTTACTAACTGTGTGGTTTGGTTGACTGGGCCAC 918

Qy     839 AGCCACGGCCACCAACCAGCAGCAGGAGGAGAACCCCAGCGTCCGGGAGCTGCCTTCATCCAT 898
           |||||| ||||||||||| ||||| || |||||||||| ||  || ||||| || ||||||
Db     919 CAAGGAAGTACAAGCCAATGCCAGCGTCAGAGCTGCAGAGAGCTGCT 969

Qy     899 GTGATCCGGCGACTTTATCGACAGCATGGGTGTCCTAGTGGCCAGCCTATATTTTATACTTC 958
           |||| || || || ||| || ||||| ||| || || |||| || || ||||| || |||
Db     959 CAAGGAAGTACAAGCCAATGCCAGCGTCAGAGCTGCAGAGAGCTGCTTAATTAGTGCACTTAT 1029

Qy     959 AAGCCACAAATACAAGTATGTAGACCCCATCTCACCTTCGTTCTTCTCCATCGGTCCTG 1018
           ||||| ||||| || ||||| ||||| || |||||| || |||||||||||  || ||
Db     970 GCCCTTGGAGATCTATTCAGAGTACTCAGTGTCCTTAATTAGTGCACTTATTCTACTTTT 1029

Qy    1019 GGGACAACCTTGACCATCCTGAGAGATGTTGATCCTGGTTGATTGGGAAGGGGACCCCCAAG 1078
           |||| || ||||||||| ||||||| |||| ||| ||||||||||| ||||| |||||||
Db    1030 AAGCCAGGATATAAAATAGCCGAGACCCAAATCTGCACATTCCATCCTGGTCTG 1089

Qy    1079 GCGGTTGACTTCACAGCTGTTCGTGATCTGCTGCTCTCCATCTTACTCATGGTTGCCAAGG 1138
           ||||| ||| ||||||||  || ||| || || || || |||||| |||||||| ||||||
Db    1090 GCCAGCACCATACTTCTTAACACTCTGTCTAACAATGAAGTAATTCTCTCAGCTCATGTTCCT 1149

Qy    1139 CACAGGCCTGCATATCTCGGGCACTGACGGTGCGCCACCTGTTCTGTGTCTGTCACACATGCC 1198
           ||||| || ||||| || ||||| ||| ||||| ||||| ||  || || || |||||||||
Db    1150 AGCCTGAATTACAGTGTGTGAAAAGAGCTTATTTAACCAGTCGACAGGGGTGTGTCCTGTG 1209

Qy    1199 ATTGCTCAGAAATACAGACGCCGGCCAGCCTGTGCTGAAGACAGCCAGCAGCGCGCTCCAAGG 1258
           |||||||||||| ||||||||||  ||||||| || ||||| || ||||| || || |||||
Db    1210 CACAGCCTGCACACTCGTCTTCTAACAATGAAGTAATTCTCTCAGCTCATGTTCCT 1269

Qy    1259 AAGTTCCACTTCACACCGTGACCTCCAGATCGAGGACTACTGCGAGGACATGAAGGAC 1318
           |||||||||||||| || ||||| || ||||| ||||||  | ||||||| || ||
Db    1270 ACAGCAGCCACCGGGACAGCCGACCCCAAGGTGGTTCGGAAAGAAATTGCTTAAAGCAAA 1329

Qy    1259 AAGTTCCACTTCACACCGTGACCTCCAGATCGAGGACTACTGCGAGGACATGAAGGAC 1318
           |||||| ||||| |||| ||| || |||||||||  ||| |||  ||||| || ||
Db    1330 AGCTTTACGATCGCACTCACTCCACCATTCAGATGGAATCTCCAGTTGACCAGGACCCGAC 1389
```

```
Qy   1319 TGTCAGGCATGCCAGGGCCCCTCAGACTGACT 1350
          ||| |||| |||||  |||| |||||||||||
Db   1390 TGCCTTTTCTGTGAAGACCCCTGTGACTAGCT 1421
```

RESULT 15

LOCUS   AX351593   2852 bp   DNA   linear   PAT 06-FEB-2002
DEFINITION   Sequence 20 from patent WO0194409.
ACCESSION   AX351593
VERSION   AX351593.1 GI:18616891
KEYWORDS   .
SOURCE   human.
  ORGANISM   Homo sapiens
             Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
             Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1
  AUTHORS   Hirst,S.K., Harlocker,S.L., Dillon,D.C. and Kalos,M.D.
  TITLE   Compositions and methods for the therapy and diagnosis of
             pancreatic cancer
  JOURNAL   Patent: WO 0194409-A 20 13-DEC-2001;
             CORIXA CORPORATION (US)
FEATURES   Location/Qualifiers
     source   1..2852
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
BASE COUNT   863 a   601 c   620 g   768 t
ORIGIN

Query Match   19.1%; Score 309.6; DB 6; Length 2852;
Best Local Similarity   57.9%; Pred. No. 8.4e-56;
Matches 574; Conservative   0; Mismatches 409; Indels   9; Gaps   1;

```
Qy    359 GAGCTGGCTGCCCAGAGCAACCATCACTCACTGCCCATGCTCAGAAGGGTCCTGACAGTCACTGT 418
          || ||||||| |||| ||| | ||| || || ||||||||||| ||||||| |  |||
Db    439 GAGCTGGAGTCAGGAGGCATGTTACCACTGCCCAAGCCCACAGAAAAGGGG 498

Qy    419 GACCCCAAGAAAGGGGAAAGGCCCAGCGCCAGTCTATGTAGCCTCTGCCATCTGCCCTGTG 478
          |||| || |||| ||||| ||  ||| ||| || || ||||| |||| |||| |||||
Db    499 GCGAAGTCAGTACGCCTATGCCAAAGTGGAAAACTTCGTTCTTGCTTCAGCAATATGCTTCATT 558

Qy    479 TTCATGATCGGAGAAGTCGTTGGGTGGGTTACCTGGCACACAGCTTGGGCTGTCATTGACTGAC 538
          |||||||  || || | | ||||||| ||||||| ||||| || ||| || ||| ||||||
Db    559 TTCATGATTGCAGAGGTCGTGGGTGGGGCACCATTGCCTGGGAGTCTTGCTGTGTTGTCACAGAT 618

Qy    539 GCAGCACACCTCACTGACTTTGCCAGCATGCTCATCAGCCTCTTCTCCCTCTGGATG 598
          |||| ||||| ||||||  ||| ||||| ||||||||||| |||||||||| ||
Db    619 GCTGCCACCCTCTTAATTGACCTGACCAGTTCCTGCTGTCTCTTCTCCCTGTGGTTG 678

Qy    599 TCCTCCCCGCCAGCCACCAAGACCATGAACTTTGGCTGGCAGAGGAGTGAGATCTTGGGA 658
          | ||| ||  |||||| || ||| |||  ||||| || | ||| || ||||||||||
Db    679 TCATCGAAGCCTCCCCTCTAAGCGGGCTGAACATTTGGATGGCACCGACGAGGAATCCTTGGT 738

Qy    659 GCCCTGGTCTTCTTGTACTGTGTCCATCTGGGTGGTCTGACGGGGGGTACTGGGCTGTG 718
          ||||||||||||| |||| |||||||| ||||| ||||||  || |||||||||| |
Db    739 GCCCTGCTCTCCATCCTGTGCATCTGGTGTGGTTGCACCAGAGAGATGGCTAGTCACCTGGCATGT 798

Qy    719 GAGCGGGCTGATCTCTGGGGACTATGAAATTGACGGGGGGACCATGCTGATCACGTCGGGC 778
          |||| || |||| | |||| |||  || || || ||| ||||| |||||||||| |||||
Db    799 GAGCGGGCTGCTGTATCCTGATTTACCAGATCCAGGCGACTGTGATGACATCGTTTCCAGC 858

Qy    779 TGGCGCTGTGGGTGTGAACATCATAAATGGGGGTTGACCCTTCACCAGTGCCCATGGGCCAC 838
          ||||| || ||  | | ||| || ||||  ||||  ||| |||||| || |||| ||||||
Db    859 TGCGCACTGGGGGGCAACATTGTACTTAACTGTGGTTTTGCACCAGAGAGATGCCTTGGCCAC 918

Qy    839 AGCCACGGCACCAACCAGGAGGAGAACCCCAGTCTGCCTTCATCCAT 898
          |||| |||||||||| ||||| || |  |||||||||||||||||||
Db    919 AATCA------CAAGGAAGTACAAGCCAATGCCACGGCTCAGAGCTGCTGTTTTGTGGCAT 969

Qy    899 GTGATCGGCGACTTTATGCAGAGCATGGGTGTGCCTAGTGGCAGCCTATATATTTTATACTTC 958
          || |  ||| ||||   |||  ||| | | | ||||| | |||| |||||| || |||||
Db    970 GCCCTTGGAGATCTATTTCAGAGTATGCACTTATTTAGTCACTTATTCTACTTT 1029
```

THIS PAGE BLANK (USPTO)

OM nucleic - nucleic search, using sw model

GenCore version 5.1.4_p5_4578
Copyright (c) 1993 - 2003 Compugen Ltd.

Run on: March 29, 2003, 08:02:28 ; Search time 441.122 Seconds
(without alignments)
8255.046 Million cell updates/sec

Title: US-09-691-219-1
Perfect score: 1617
Sequence: 1 cgcggtgaacacgcaggg..........aaaaaaaaaaaaatgtt 1617
Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0
Searched: 2185239 seqs, 1125999159 residues
Total number of hits satisfying chosen parameters: 4370478

Post-processing: Minimum Match 0%
Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Listing first 45 summaries

Database : N_Geneseq_101002:*
```
 1: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1980.DAT:*
 2: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1981.DAT:*
 3: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1982.DAT:*
 4: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1983.DAT:*
 5: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1984.DAT:*
 6: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1985.DAT:*
 7: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1986.DAT:*
 8: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1987.DAT:*
 9: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1988.DAT:*
10: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1989.DAT:*
11: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1990.DAT:*
12: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1991.DAT:*
13: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1992.DAT:*
14: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1993.DAT:*
15: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1994.DAT:*
16: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1995.DAT:*
17: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1996.DAT:*
18: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1997.DAT:*
19: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1998.DAT:*
20: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1999.DAT:*
21: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA2000.DAT:*
22: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA2001A.DAT:*
23: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA2001B.DAT:*
24: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA2002.DAT:*
```

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	1617	100.0	1617	24	ABN83946	Human transporter
2	1268	78.4	1280	24	AAD36301	Human transporter
3	601	37.2	2291	22	AAH16506	Human cDNA sequence
4	590	36.5	732	22	AAH07440	Human cDNA clone (
5	521	32.2	783	22	AAH07460	Human cDNA clone (
6	521	32.2	1440	22	AAH16516	Human cDNA sequenc
7	363.6	22.5	11101	24	ABN83947	Human transporter
8	309.6	19.1	1318	24	ABK71914	Human transporter
9	309.6	19.1	1431	24	ABK71912	Human cDNA encodin
10	309.6	19.1	2165	24	ABK88011	cDNA encoding huma
11	309.6	19.1	2297	23	ABV23919	Human prostate exp
12	309.6	19.1	2297	23	ABV29801	Human prostate exp
13	309.6	19.1	2823	22	AAF27714	Human transport pr
14	267.8	16.6	2852	22	AAD28259	Human pancreatic t
15	267.8	16.6	1623	24	ABK71913	Human cDNA encodin
16	175.6	10.9	1290	24	ABK92126	prostate cancer-as
17	175.6	10.9	1826	23	AAD38825	Human PSNA cDNA, P
18	167	10.3	1806	23	ABL07015	Drosophila melanog
19	165.6	10.2	5034	23	ABL26319	Drosophila melanog
20	163	10.1	2249	23	ABV24622	Human prostate exp
21	152.6	9.4	10793	23	ABL26318	Human prostate exp
22	144.6	8.9	779	22	AAL21741	Drosophila melanog
23	143.8	8.9	1551	22	AAL16065	Maize zinc transpo
24	141.6	8.8	292	22	AAL13518	Human breast cance
25	141.6	8.8	292	22	AAL21864	Human breast cance
26	140	8.7	389	22	AAL17611	Human breast cance
27	137.8	8.5	280	22	AAL13516	Human breast cance
28	137.4	8.5	285	22	AAL22385	Human breast cance
29	136.4	8.4	414	22	AAL10536	Human breast cance
30	136.4	8.4	434	22	AAL17999	Human breast cance
31	134.6	8.3	449	22	AAL10896	Human breast cance
32	134.6	8.3	1873	22	AAD16066	Rice zinc transpor
33	132	8.2	408	22	AAL22870	Wheat zinc transpo
34	131.8	8.2	790	22	AAL16068	Human breast cance
35	118.2	7.3	288	22	AAL12994	Human breast cance
36	115.6	7.1	1368	21	AAC39571	Arabidopsis thalia
37	113.8	7.0	1372	22	AAC47581	Arabidopsis thalia
38	113.8	7.0	1208	22	AAD16067	Soybean zinc trans
39	106.2	6.6	1763	23	ABL17389	Drosophila melanog
40	99	6.1	21109	23	ABL07014	Drosophila melanog
41	93	5.8	171	24	ABN23701	Human ORFX polynuc
42	89.4	5.5	171	23	ABV36096	Human prostate exp
43	87.8	5.4	478	23	ABV45153	Human prostate exp
44	87.8	5.4	193	23	ABV15301	Human prostate exp
45	87.4	5.4	193	23		Human prostate exp

ALIGNMENTS

```
RESULT 1
ABN83946

ID   ABN83946 standard; cDNA; 1617 BP.
XX
AC   ABN83946;
XX
DT   02-OCT-2002  (first entry)
XX
DE   Human transporter protein encoding cDNA.
XX
KW   Human; transporter protein; zinc transporter; pharmacogenomic analysis;
KW   diagnosis; drug screening; gene therapy; kidney; testis; heart; placenta;
KW   small intestine; liver; chromosome 1; gene; ss.
XX
OS   Homo sapiens.
XX
FH   Key            Location/Qualifiers
FT   CDS            230..1348
FT                  /tag= a
FT                  /product= "transporter protein"
XX
PN   WO200224910-A2.
XX
PD   28-MAR-2002.
XX
PF   20-SEP-2001; 2001WO-US292218.
XX
PR   20-SEP-2000; 2000US-234160P.
PR   19-OCT-2000; 2000US-0691219.
XX
PA   (PEKE ) PE CORP NY.
```

PI   Wei M, Ketchum KA, Di Francesco V, Beasley EM;
XX
DR   WPI; 2002-404954/43.
DR   P-PSDB; ABB83084.
XX
PT   Novel human transporter proteins, related to zinc transporter
PT   subfamily, useful as model for developing human therapeutic targets and
PT   serves as target for human therapeutics -
XX
PS   Claim 4(b); Fig 1; 75pp; English.
XX
CC   The invention relates to an isolated human transporter protein that is
CC   related to the zinc transporter subfamily. Polynucleotides and
CC   polypeptides of the invention are useful for treating a disease or
CC   condition mediated by human transporter protein. The proteins also
CC   provide a target for diagnosing a disease or predisposition to disease
CC   mediated by the peptide, and in pharmacogenomic analysis. The peptides
CC   are also useful for treating a disorders characterised by absence of,
CC   inappropriate or unwanted expression of the protein. The nucleic acids
CC   are also useful in drug screening assays and as a target for treatment
CC   by the compounds identified through drug screening. The invention also
CC   provides vectors for gene therapy in patients with aberrant expression
CC   of the gene encoding the transporter protein. The gene of the invention
CC   has been found to be expressed in humans in the kidney, testis, heart,
CC   placenta, small intestine and liver. The gene has been localised to
CC   human chromosome 1. The current sequence represents the human transporter
CC   protein encoding cDNA.
XX
SQ   Sequence 1617 BP; 351 A; 473 C; 489 G; 304 T; 0 other;

Query Match   100.0%; Score 1617; DB 24; Length 1617;
Best Local Similarity   100.0%;
Matches 1617; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy   1 CGCGGTGAACAGCACCAGAGGGGAGCCGCCTTGGCACGCGGAGCCGAGCTGCAGCC 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1 CGCGGTGAACAGCACCAGAGGGGAGCCGCCTTGGCACGCGGAGCCGAGCTGCAGCC 60

Qy   61 GCAGCGGGGAGACCGGGGGAGCTCAGGGGCCCGAGGGAGTGAGCCACCTCG 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||
Db   61 GCAGCGGGGAGACCGGGGGAGCTCAGGGGCCCGAGGGAGTGAGCCACCTCG 120

Qy   121 CGGGGCCCTCGGGACGTGGGGTGACCGCCACCTGCCCAACTTTCAGGGC 180
              |||||||||||||||||||||||||||||||||||||||||||||||||
Db   121 CGGGGCCCTCGGGACGTGGGGTGACCGCCACCTGCCCAACTTTCAGGGC 180

Qy   181 GCACTCGGCGCGGGGGCGGCTGCCGGCGACTCGGCGGACTGATGGCCAA 240
              |||||||||||||||||||||||||||||||||||||||||||||||||
Db   181 GCACTCGGCGCGGGGGCGGCTGCCGGCGACTCGGCGGACTGATGGCCAA 240

Qy   241 GGAGAAGCAGCATCTGTTGGACGCGCAATTCCGGTCATCACACGGAATCTCTGTG 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   241 GGAGAAGCAGCATCTGTTGGACGCGCAATTCCGGTCATCACACGGAATCTCTGTG 300

Qy   301 GCAGGAAGGGGCTGGCTGGATTCCTCTGCCCCGACCTGGACTTGCAGGCCAATTGA 360
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   301 GCAGGAAGGGGCTGGCTGGATTCCTCTGCCCCGACCTGGACTTGCAGGCCAATTGA 360

Qy   361 GCTGGCTGCCCAGAGCAAACATCACTGCCATGCTCAGAAGGGTCCCCACTCACTGTGA 420
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   361 GCTGGCTGCCCAGAGCAAACATCACTGCCATGCTCAGAAGGGTCCCCACTCACTGTGA 420

Qy   421 CCCCAAGAAGGGGAAGCCCAGCGCCGCTGTATGTAGCCTCTGCCATCTGCCTGTTGTT 480
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   421 CCCCAAGAAGGGGAAGCCCAGCGCCGCTGTATGTAGCCTCTGCCATCTGCCTGTTGTT 480

Qy   481 CATGATCGGAGAAGTCGTTGGTGGGTACCTGGCACACAGCTTGGCTGTCATGACTGACGC 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   481 CATGATCGGAGAAGTCGTTGGTGGGTACCTGGCACACAGCTTGGCTGTCATGACTGACGC 540

Qy   541 AGCACACCTGCTCACTGACTTTGCCAGCATGCTCATCAGCCTCTTCTCCCCTGGAGTGC 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   541 AGCACACCTGCTCACTGACTTTGCCAGCATGCTCATCAGCCTCTTCTCCCCTGGAGTGC 600

Qy   601 CTCCCGGCCCAGCCACCAGAACCATGAACTTTGCTGTGCAGAGAGAGACTGAGATCTTGGAGACC 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   601 CTCCCGGCCCAGCCACCAGAACCATGAACTTTGCTGTGCAGAGAGAGACTGAGATCTTGGAGACC 660

Qy   661 CCTGGGTCTCGTACTGTCCATCTGGGTGCGTGACGGGGGTACTGGTGTCGTGTGGA 720
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   661 CCTGGGTCTCGTACTGTCCATCTGGGTGCGTGACGGGGGTACTGGTGTCGTGTGGA 720

Qy   721 GCGGCCTGATCTCTCGGGAGGTCATATGAAAATTGACGGGGGGGGACCATGTCTGATCACGTGCGGGGGCTG 780
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   721 GCGGCCTGATCTCTCGGGAGGTCATATGAAAATTGACGGGGGGGGACCATGTCTGATCACGTGCGGGGGCTG 780

Qy   781 CGCTGTGCCTGTGTGAACATCATAATGGGGGTTGACCCTTCACCAGTCTGGCCATGGGCACAG 840
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   781 CGCTGTGCCTGTGTGAACATCATAATGGGGGTTGACCCTTCACCAGTCTGGCCATGGGCACAG 840

Qy   841 CCACGGCACCACCAACCAGCAGGAGGAGGAACCCCAGCGTCCGAGCTGCCTTCATCCATGT 900
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   841 CCACGGCACCACCAACCAGCAGGAGGAGGAACCCCAGCGTCCGAGCTGCCTTCATCCATGT 900

Qy   901 GATCGGGCGACTTTATGCAGAGCATGGGGTGTCCTAGTGCGCAGCCTATATTTTATACTTCAA 960
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   901 GATCGGGCGACTTTATGCAGAGCATGGGGTGTCCTAGTGCGCAGCCTATATTTTATACTTCAA 960

Qy   961 GCCAGAAATACAAGTATGTPAGACCCCATCCTGCACCTTCGTCTTCTCCCATCCCTTGTCCTGGG 1020
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   961 GCCAGAAATACAAGTATGTPAGACCCCATCCTGCACCTTCGTCTTCTCCCATCCCTTGTCCTGGG 1020

Qy   1021 GACAAACCTTGACCCATCCTGAGAGAGATGTGATCCTGGTGTTGGATTGGAGGAGGACACCCCAAGGG 1080
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1021 GACAAACCTTGACCCATCCTGAGAGAGATGTGATCCTGGTGTTGGATTGGAGGAGGACACCCCAAGGG 1080

Qy   1081 CGTTGACTTCACAGCTGTTCGTGATCTGCGTCGTGGGTTCGGTGGAGAGGGGTGAGAAGCCCTGCA 1140
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1081 CGTTGACTTCACAGCTGTTCGTGATCTGCGTCGTGGGTTCGGTGGAGAGGGGTGAGAAGCCCTGCA 1140

Qy   1141 CAGCCTGCATATCTGGGCACTGACGGTGGCCAGCCGTTTCTGTCTCTGTGCCACATCGCCAT 1200
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1141 CAGCCTGCATATCTGGGCACTGACGGTGGCCAGCCGTTTCTGTCTCTGTGCCACATCGCCAT 1200

Qy   1201 TGCTCAGAATACAGAGACGCCCCAGGCTGTCGTGAAGACACGCCGCCTCCAAGGGAA 1260
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1201 TGCTCAGAATACAGAGACGCCCCAGGCTGTCGTGAAGACACGCCGCCTCCAAGGGAA 1260

Qy   1261 GTTCCACTTCCACACCGTGACCATCCAGATCGAGGAGGACTACTCGGAGGAGGACATGAAGGACTG 1320
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1261 GTTCCACTTCCACACCGTGACCATCCAGATCGAGGAGGACTACTCGGAGGAGGACATGAAGGACTG 1320

Qy   1321 TCAGGCAATGCCAGGGCCCCTCAGACTGACTGCTGCAGCACCAACTGGGGCATGAAC 1380
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1321 TCAGGCAATGCCAGGGCCCCTCAGACTGACTGCTGCAGCACCAACTGGGGCATGAAC 1380

Qy   1381 AGGACCCTGCAGGTGGCTGGACTGTGGTTGTCCCCAGGCGCCAGGACTTTGCCTACCCC 1440
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1381 AGGACCCTGCAGGTGGCTGGACTGTGGTTGTCCCCAGGCGCCAGGACTTTGCCTACCCC 1440

Qy   1441 AGCTGTGTTATAAACCAGGTCCCCTCCTGACCTCCTGCCCCACTCCAGGAAATGGAGCTCT 1500
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1441 AGCTGTGTTATAAACCAGGTCCCCTCCTGACCTCCTGCCCCACTCCAGGAAATGGAGCTCT 1500

Qy   1501 TCCCAGCCTCCCATCTGACTACAGCCAGGGGACTCAGCGGGGTATAAAAGCTAGTGTG 1560
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1501 TCCCAGCCTCCCATCTGACTACAGCCAGGGGACTCAGCGGGGTATAAAAGCTAGTGTG 1560

Qy   1561 ACCCTGAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAATGTT 1617
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1561 ACCCTGAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAATGTT 1617

RESULT 2
AAD36301
ID   AAD36301 standard; cDNA; 1280 BP.

```
XX
AC    AAD36301;
XX
DT    09-AUG-2002 (first entry)
XX
DE    Human transporter and ion channel (TRICH) 4 cDNA.
XX
KW    Human; transporter and ion channel; TRICH; transport disorder;
KW    diabetes mellitus; angina; Alzheimer's disease; neurological; epilepsy;
KW    stroke; Huntington's disease; meningitis; muscle; myocarditis; cancer;
KW    infectious myositis; arrhythmia; asthma; immunological; gene therapy;
KW    acquired immunodeficiency syndrome; AIDS; allergy; atherosclerosis;
KW    cell proliferative disorder; cerebroprotective; cirrhosis; hepatitis;
KW    transgenic; neuroprotective; anticonvulsant; nootropic; cytostatic;
KW    antiinflammatory; hepatotropic; psoriasis; gene; ss.
OS    Homo sapiens.
XX
FH    Key             Location/Qualifiers
FT
FT    CDS             144..1262
FT                    /*tag= a
FT                    /product= "Human TRICH4 protein"
XX
PN    WO200222684-A2.
XX
PD    21-MAR-2002.
XX
PF    14-SEP-2001; 2001WO-US28938.
XX
PR    15-SEP-2000; 2000US-232685P.
PR    22-SEP-2000; 2000US-234842P.
PR    29-SEP-2000; 2000US-236882P.
PR    05-OCT-2000; 2000US-239057P.
PR    13-OCT-2000; 2000US-240540P.
PR    18-OCT-2000; 2000US-241700P.
XX
PA    (INCY-) INCYTE GENOMICS INC.
XX
PI    Lee EA, Yue H, Lal PG, Walia NK, Baughn MR, Warren BA, Lee S;
PI    Sanjanwala MS, Yao MG, Ramkumar J, Thornton M, Gandhi AR;
PI    Policky JL, Elliott VS, Arvizu C, Raumann BE, Bruns CM, Naini A;
PI    Hafalia AJA, Nguyen DB, Xu Y, Lu DAM, Ison CH, Griffin JA;
PI    Reddy RM, Burford N;
XX
DR    WPI; 2002-393948/42.
DR    P-PSDB; AAE22905.
XX
PT    Polypeptides of human transporters and ion channels, useful for
PT    diagnosing, treating or preventing transport, neurological, muscle,
PT    immunological and cell proliferative disorders -
XX
PS    Claim 12; Page 184; 204pp; English.
XX
CC    The invention relates to human transporters and ion channels (TRICH)
CC    and their corresponding nucleic acid sequences. TRICH is useful for
CC    screening an agonist/antagonist that modulates its activity. TRICH is
CC    useful as an immunogen for preparing antibodies which are useful for
CC    diagnosing a condition of disease associated with its expression in a
CC    subject, and for detecting and purifying it from a sample. TRICH DNA
CC    is useful as probe or a primer for assessing toxicity of a test
CC    compound. Composition comprising TRICH or its agonist is useful for
CC    treating a disease or condition associated with decreased expression
CC    of functional TRICH and composition comprising TRICH antagonist is
CC    useful for treating a disease or condition associated with TRICH
CC    overexpression of TRICH sequence is used in the diagnosis and
CC    treatment of transport disorder e.g. diabetes mellitus, angina,
CC    Alzheimer's disease; neurological disorder e.g. epilepsy, stroke,
CC    Huntington's disease; infectious myositis, arrhythmias, asthma,
CC    e.g. myocarditis, bacterial and viral meningitis, muscle disorder
CC    immunological disorder e.g. acquired immunodeficiency syndrome (AIDS),
CC    allergies, atherosclerosis; and cell proliferative disorders e.g.
CC    cirrhosis, hepatitis, psoriasis and cancers. TRICH DNA is used in
CC    gene therapy. TRICH DNA is useful for creating knockin humanised
CC    animals (pigs) or transgenic animals (mice or rats) to model human
CC    disease. The present sequence is human TRICH cDNA.
XX
SQ    Sequence 1280 BP; 248 A; 383 C; 392 G; 257 T; 0 other;

      Query Match      78.4%;   Score 1268;  DB 24;   Length 1280;
      Best Local Similarity   99.9%;   Pred. No. 7.2e-250;
      Matches 1279;  Conservative   0;  Mismatches   0;   Indels   1;   Gaps   1;
```

```
Qy        88  CCGGAGGAGCCGGGCCGGGAGTGAGGCCGCCACCTCGCGGGG-CCCTCGGGGCGAGGTGGGGTGGGGGTGGGGTGAG   146
              ||||||||||||||||||||||||||||||||||||||||||||  ||||||||||||||||||||||||||||||||
Db         1  CCCGAGGAGCCGCGGGTGGGCCGGGGAGTGAGGCGACCACCTCGCGGTCCCTCCTTCCCTCGGGCACTGCGAGGTGGGGAG    60

Qy       147  CGCCACCCCGGAGTCCCGCGCGCAACTTTCAGGGGCGGCCACTCGGGCCGCGGGCGGGGGCT   206
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        61  CGCCACCCCGGAGTCCCGCGCAACTTTCAGGGGCGGCCACTCGGGCCGCGGGCGGGGGCT   120

Qy       207  GCCCGGGACTCGGGCCGGGGACTGCCATGGGAGGCCAAGGAGGAGCAGCAATCTGTTGGACGCCA   266
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       121  GCCCGGGACTCGGGCCGGGGACTGCCATGGGAGGCCAAGGAGGAGCAGCAATCTGTTGGACGCCA   180

Qy       267  GGGCGGGGCAATCCGGTCATACACGGGGAATCCTCTGTGCGCAGAAGGGGCTGGCTGGATTCCTC   326
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       181  GGGCGGGGCAATCCGGTCATACACGGGGAATCCTCTGTGCGCAGAAGGGGCTGGCTGGATTCCTC   240

Qy       327  TGCCCCGGACCTGGGCCTGGACTTGCGAGGCCATTGAGCTGGCTGGCTGCCGGAGAGCAACCATCACT   386
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       241  TGCCCCGGACCTGGGCCTGGACTTGCAGGCCATTGAGCTGGCTGGCTGCCAAGGAGAGCAACCATCACT   300

Qy       387  GCCATGCTCAGAAGGGTCCTTGACAGTCACTGTGACCCCAAGAAGGGAAGGCCAGGCGC   446
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       301  GCCATGCTCAGAAGGGTCCTTGACAGTCACTGTGACCCCAAGAAGGGAAGGCCAGGCGC   360

Qy       447  AGCGTGTATGTTAGCCTTCTGCCATCCTGCCATCCTGTGTTGTTCATGATCGGAGAAGTGTTTGGGT   506
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       361  AGCGTGTATGTTAGCCTTCTGCCATCCTGCCATCCTGTGTTGTTCATGATCGGAGAAGTGTTTGGGT   420

Qy       507  ACCTGGCCACACAGCTTGGCTGTCGTCATGACTGACGAGCAACCTGCTCACTGACTTTGCCA   566
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       421  ACCTGGCCACACAGCTTGGCTGTCGTCATGACTGACGAGCAACACCTGCTCACTGACTTTGCCA   480

Qy       567  GCAATCCTCATCAGCTCTTTCTCCCCTCTGATGTTCCCGGCCCCAGCCACCAAGACCATGA   626
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       481  GCATGCTCATCAGCCTCTTTCTCCCCTCTGATGTTCCCGGCCCCAGCCACCAAGACCATGA   540

Qy       627  ACTTTGGCTGGCAGAGAGCTGAGATCTTGGGAGTCTTGGTGTCTCTGTTACTGTCCATCTGGG   686
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       541  ACTTTGGCTGGCAGAGAGCTGAGATCTTGGGAGTCTTGGTGTCTCTGTTACTGTCCATCTGGG   600

Qy       687  TCGTGACGGGGGGTACTGGTGTACCTGGTCTGGAGCGGGCTGATCTCTGGGGACTATGAAA   746
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       601  TCGTGACGGGGGGTACTGGTGTACCTGGTCTGGAGCGGGCTGATCTCTGGGGACTATGAAA   660

Qy       747  TTGACAGGGGGGACCATGGTCATCACGTCCGGCTCGCGCTCTGGCTGTGGACATCATAATGG   806
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       661  TTGACGGGGGGGACCATGGTCATCACGTCCGGCTCGCGCTGTGGCTGTGGACATCATAATGG   720

Qy       807  GGTTGACCCTTCACCAGTCTGGCCATGGGCACAGCCACCACCAACCAGCAGGAGG   866
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       721  GGTTGACCCTTCACCAGTCTGGCCATGGGCACAGCCACCACCAACCAGCAGGAGG   780

Qy       867  AGAACCCCAGCGTCCGAGCTGCCTTCATCCATGTGATCGGCGACTTTAGTGCAAGGCCATGG   926
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       781  AGAACCCCAGCGTCCGAGCTGCCTTCATCCATGTGATCGGCGACTTTATTGCAGAGGCATGG   840

Qy       927  GTGTCCTTACGTGGCAGCCCTATATTTTATACTTCAAGCAGCAGAATACAAGTATGTAGACCCCA   986
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       841  GTGTCCTAGTGGCAGCCCTATATTTTTATACTTCAAGCCAGAATACAAGTATGTAGACCCCA   900

Qy       987  TCTGCACCTTCGTCGTCTTCTCCCATCCTGGGTCCTGGGGGGACAACCTTGACCATCCTGAGAGATG   1046
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       901  TCTGCACCTTCGTCGTCTTCTCCCATCCTGGGTCCTGGGGGGACAACCTTGACCATCCTGAGAGATG   960
```

```
Qy    1047 TGATCCTGGTGTTGATGGAAGGAGACCCCAAGGGCGGTTGACTTCACAGCTGTTCGTGATC 1106
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     961 TGATCCTGGTGTTGATGGAAGGAGACCCCAAGGGCGGTTGACTTCACAGCTGTTCGTGATC 1020
XX
Qy    1107 TGCTGCTGTCGGTGGAGGGGGTAGAAGCCCTGCACAGCCTGCATANTCTGGGCACTGACGG 1166
           ||||||||||||||||||||||||||||||||||||| ||||||||||||||||||||||
Db    1021 TGCTGCTGTCGGTGGAGGGGGTAGAAGCCCTGCACAGCCTGCATATTCTGGGCACTGACGG 1080
XX
Qy    1167 TGGGCCAGCCTGTTCTGTCTGTGTCNCACATCNCGAGAANTACAGACGCGCCAGGCTG 1226
           ||||||||||||||||||||||||| |||||||| ||||| |||||||||||||||||||
Db    1081 TGGGCCAGCCTGTTCTGTCTGTGTCCACATCCCCAGAATACAGACGCGCCAGGCTG 1140
XX
Qy    1227 TGCTGAAGACAGCAGCCAGCACGCCGCCGCCTCCAAGGGAAGTTCCACTTCCACACCGTGACCATCC 1286
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1141 TGCTGAAGACAGCAGCCAGCACGCCGCCGCCTCCAAGGGAAGTTCCACTTCCACACCGTGACCATCC 1200
XX
Qy    1287 AGAATGGAGGACTACTCGGAGGACATGTCAGGACTCAGGGAGACTGTCAGGACT 1346
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1201 AGAATGGAGGACTACTCGGAGGACATGAAGGACTGTGCAGGGCCCCCCTCCAAGGGCGTTG 1260
XX
Qy    1347 GACTGCTCAGCCAGGCACCA 1366
           ||||||||||||||||||||
Db    1261 GACTGCTCAGCCAGGCACCA 1280
```

```
RESULT 3
AAH16506
ID     AAH16506 standard; cDNA; 2291 BP.
XX
AC     AAH16506;
XX
DT     26-JUN-2001 (first entry)
XX
DE     Human cDNA sequence SEQ ID NO:15542.
XX
KW     Human; primer; detection; diagnosis; antisense therapy; gene therapy; ss.
OS     Homo sapiens.
XX
PN     EP1074617-A2.
XX
PD     07-FEB-2001.
XX
PF     28-JUL-2000; 2000EP-0116126.
XX
PR     29-JUL-1999; 99JP-0248036.
PR     27-AUG-1999; 99JP-0300253.
PR     11-JAN-2000; 2000JP-0118776.
PR     02-MAY-2000; 2000JP-0183767.
PR     09-JUN-2000; 2000JP-0241899.
XX
PA     (HELI-) HELIX RES INST.
XX
PI     Ota T, Isogai T, Nishikawa T, Hayashi K, Saito K, Yamamoto J;
PI     Ishii S, Sugiyama T, Wakamatsu A, Nagai K, Otsuki T;
XX
DR     WPI; 2001-318749/34.
XX
PT     Primer sets for synthesizing polynucleotides, particularly the 5602
PT     full-length cDNAs defined in the specification, and for the detection
PT     and/or diagnosis of the abnormality of the proteins encoded by the
PT     full-length cDNAs -
XX
PS     Claim 8; SEQ ID 15542; 2537pp + CD ROM; English.
XX
CC     The present invention describes primer sets for synthesising 5602
CC     full-length cDNAs defined in the specification. Where a primer set
CC     comprises: (a) an oligo-dT primer and an oligonucleotide complementary
CC     to the complementary strand of a polynucleotide which comprises one of
CC     the 5602 nucleotide sequences defined in the specification, where the
CC     oligonucleotide comprises at least 15 nucleotides; or (b) a combination
CC     of an oligonucleotide comprising a sequence complementary to the
CC     complementary strand of a polynucleotide which comprises a 5'-end
```

```
CC     sequence and an oligonucleotide comprising a sequence complementary to a
CC     polynucleotide which comprises a 3'-end sequence, where the
CC     oligonucleotide comprises at least 15 nucleotides and the combination of
CC     the 5'-end sequence/3'-end sequence is selected from those defined in
CC     the specification. The primer sets can be used in antisense therapy and
CC     in gene therapy. The primers are useful for synthesising polynucleotides,
CC     particularly full-length cDNAs. The primers are also useful for the
CC     detection and/or diagnosis of the abnormality of the proteins encoded by
CC     the full-length cDNAs. The primers allow obtaining of the full-length
CC     cDNAs easily without any specialised methods. AAH03166 to AAH13628 and
CC     AAH13633 to AAH18742 represent human cDNA sequences; AAB92446 to
CC     AAB95893 represent human amino acid sequences; and AAH13629 to AAH13632
CC     represent oligonucleotides, all of which are used in the exemplification
CC     of the present invention.
XX
SQ     Sequence 2291 BP; 474 A; 663 C; 647 G; 507 T; 0 other;
```

```
Query Match             37.2%;  Score 601;  DB 22;  Length 2291;
Best Local Similarity   100.0%;  Pred. No. 1.9e-113;
Matches  601;  Conservative  0;  Mismatches  0;  Indels  0;  Gaps  0;

Qy     966 AATACAAGTATGTAGAACCCCATCTGACCACTTCGTCTTCTCCATCCTGGTCCTGGGGACAA 1025
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1 AATACAAGTATGTAGAACCCCATCTGACCACTTCGTCTTCTCCATCCTGGTCCTGGGGACAA 60
XX
Qy    1026 CCTTGACCCATCCTGAGAGAATGTGATCCTTGGTGTTGATGGAAGGAGACCCCAAGGGCGTTG 1085
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      61 CCTTGACCCATCCTGAGAGAATGTGATCCTTGGTGTTGATGGAAGGAGACCCCAAGGGCGTTG 120
XX
Qy    1086 ACTTCACAGCTGTTCGTGATCTGCTGCTGTCGGTGGAGGGGGTAGAAGCCCTGCACAGCC 1145
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     121 ACTTCACAGCTGTTCGTGATCTGCTGCTGTCGGTGGAGGGGGTAGAAGCCCTGCACAGCC 180
XX
Qy    1146 TGCATATCTGGGCACTGACGGTGGGCCAGCCTGTTCTGTCTGTGTCCCCATTGCTC 1205
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     181 TGCATATCTGGGCACTGACGGTGGGCCAGCCTGTTCTGTCTGTGTCCCCATTGCTC 240
XX
Qy    1206 AGAATACAGACGCGCCAGGCTGTGCTGAAGACAGCAGCCAGCACGCCGCCTCCAAGGGAAGTTCC 1265
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     241 AGAATACAGACGCGCCAGGCTGTGCTGAAGACAGCAGCCAGCACGCCGCCTCCAAGGGAAGTTCC 300
XX
Qy    1266 ACTTCCACACCGTGACCATCCAGATCAGATTACTCGGAGGACATGAAGGACTGTCAGG 1325
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     301 ACTTCCACACCGTGACCATCCAGATCAGATTACTCGGAGGACATGAAGGACTGTCAGG 360
XX
Qy    1326 CATGCCAGGCGCCCCTCAGACTGACCTGCTTCAGCGACCAGGACAGGAC 1385
           |||||||||||||||||||||||||||||||||||||||||||||||||
Db     361 CATGCCAGGCGCCCCTCAGACTGACCTGCTTCAGCGACCAGGACAGGAC 420
XX
Qy    1386 CTGCAGGTGGCTGGACTGAGTGTGTCCCCAGGGCCCAGGACTTTGCCTACCCCAGCTG 1445
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     421 CTGCAGGTGGCTGGACTGAGTGTGTCCCCAGGGCCCAGGACTTTGCCTACCCCAGCTG 480
XX
Qy    1446 TGTTATAAACCAGGTCCCCTTCTGCCCCACTCCAGGAATGGAGCTCTTCCCA 1505
           |||||||||||||||||||||||||||||||||||||||||||||||||||
Db     481 TGTTATAAACCAGGTCCCCTTCTGCCCCACTCCAGGAATGGAGCTCTTCCCA 540
XX
Qy    1506 GCCTCCCATCTGACTACAGCCACAGGTGGGGACTCAGCGGGTATAAAAGCTAGTGTGACCCT 1565
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     541 GCCTCCCATCTGACTACAGCCACAGGTGGGGACTCAGCGGGTATAAAAGCTAGTGTGACCCT 600
XX
Qy    1566 G 1566
           |
Db     601 G 601
```

```
RESULT 4
AAH07440
ID     AAH07440 standard; cDNA; 732 BP.
XX
AC     AAH07440;
XX
DT     26-JUN-2001 (first entry)
```

XX
DE Human cDNA clone (5'-primer) SEQ ID NO:4275.
XX
KW Human; primer; detection; diagnosis; antisense therapy; gene therapy; ss.
XX
OS Homo sapiens.
XX
PN EP1074617-A2.
XX
PD 07-FEB-2001.
XX
PF 28-JUL-2000; 2000EP-0116126.
XX
PR 29-JUL-1999;   99JP-0248036.
PR 27-AUG-1999;   99JP-0300253.
PR 11-JAN-2000;   2000JP-0118776.
PR 02-MAY-2000;   2000JP-0183767.
PR 09-JUN-2000;   2000JP-0241899.
XX
PA (HELI-) HELIX RES INST.
XX
PS Ota T, Isogai T, Nishikawa T, Hayashi K, Saito K, Yamamoto J;
PI Ishii S, Sugiyama T, Wakamatsu A, Nagai K, Otsuki T;
XX
DR WPI; 2001-318749/34.
XX
PT Primer sets for synthesizing polynucleotides, particularly the 5602
PT full-length cDNAs defined in the specification, and for the detection
PT and/or diagnosis of the abnormality of the proteins encoded by the
PT full-length cDNAs -
XX
CC Claim 1; SEQ ID 4275; 2537pp + CD ROM; English.
XX
CC The present invention describes primer sets for synthesizing 5602
CC full-length cDNAs defined in the specification. Where a primer set
CC comprises: (a) an oligo-dT primer and an oligonucleotide complementary
CC to the complementary strand of a polynucleotide which comprises one of
CC the 5602 nucleotide sequences defined in the specification, where the
CC oligonucleotide comprises at least 15 nucleotides; or (b) a combination
CC of an oligonucleotide comprising a sequence complementary to the
CC complementary strand of a polynucleotide which comprises a 5'-end
CC sequence and an oligonucleotide comprising a sequence complementary to a
CC polynucleotide which comprises a 3'-end sequence, where the
CC oligonucleotide comprises at least 15 nucleotides and the combination of
CC the 5'-end sequence/3'-end sequence is selected from those defined in
CC the specification. The primer sets can be used in antisense therapy and
CC in gene therapy. The primers are useful for synthesising polynucleotides,
CC particularly full-length cDNAs. The primers are also useful for the
CC detection and/or diagnosis of the abnormality of the proteins encoded by
CC the full-length cDNAs. The primers allow obtaining of the full-length
CC cDNAs easily without any specialised methods. AAH03166 to AAH13628 and
CC AAH13633 to AAH18742 represent human amino acid sequences; AAB92446 to
CC AAB95893 represent human cDNA sequences; and AAH13629 to AAH13632
CC represent oligonucleotides, all of which are used in the exemplification
CC of the present invention.
XX
SQ Sequence 732 BP; 155 A; 226 C; 200 G; 146 T; 5 other;

Query Match   36.58%; Score 590; DB 22; Length 732;
Best Local Similarity   99.8%;
Matches 601; Conservative   0; Mismatches   0; Indels   1; Gaps   1;

Qy   966 AATACAAGTATGTAGACCCCATCTGCACCTTCGTCTCTTCCAATCCTGGTCCTGGGACAA 1025
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1 AATACAAGTATGTAGACCCCATCTGCACCTTCGTCTCTTCCAATCCTGGTCCTGGGACAA 60

Qy   1026 CCTTGACCATCTCGAGAGATGTGATCTCTGGTTGGTTGATGGAAGGAGAGGGACCCCCAAGGGGCGTTG 1085
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   61 CCTTGACCATCTCGAGAGATGTGATCTCCTGGTTGATGGAAGGAGAGGGACCCCCAAGGGGCGTTG 120

Qy   1086 ACTTCACAGCTGTTCCGTGATCTGCTGCTGCTCGGTGGAGGGGGTAGAAAGCCCTGCACAGCC 1145
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   121 ACTTCACAGCTGTTCCGTGATCTGCTGCTGCTCGGTGGAGGGGGTAGAAAGCCCTGCACAGCC 180

Qy   1146 TGCATATCTGGGCACTGACGGTGGGCCCACCGTGTTCTGTCGTCGTCCACAATCGCCAATTGCTC 1205
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   181 TGCATATCTGGGCACTGACGGTGGGCCCACCGTGTTCTGTCGTCGTCCACAATCGCCAATTGCTC 240

Qy   1206 AGAATACAGACGCGCCCAGGCGTGCTGTGCTGAAGAGAGCAGCAGCCGCCGCCTCCAAGGGAAGTCC 1265
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   241 AGAATACAGACGCGCCCAGGCGTGCTGTGCTGAAGAGAGCAGCAGCCGCCGCCTCCAAGGGAAGTCC 300

Qy   1266 ACTTCCACACCGTGACCCATCCAGATCCAGATCGAGGACTACTCGGAGGAGACATGAGGGACTGTCAGG 1325
               |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   301 ACTTCCACACCGTGACCCATCCAGATCCAGATCGAGGACTACTCGGAGGAGACATGAGGGACTGTCAGG 360

Qy   1326 CATGCCAGGGGCCCCTCAGACTGACTGCTCAGCGACCACCAGGGCGGCCGATGAACAGGAC 1385
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   361 CATGCCAGGGGCCCCTCAGACTGACTGCTCAGCGACCACCAGGGCGGCCGATGAACAGGAC 420

Qy   1386 CTGCAGGTGGCGTGACTGAGTGTGCCCCAGGGCACCAGGACTTTGCCTACCCCAGCTG 1445
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   421 CTGCAGGTGGCGTGACTGAGTGTGCCCCAGGGCACCAGGACTTTGCCTACCCCAGCTG 480

Qy   1446 TGTTATAAACCAGGTGCCCCTCCTGCCCACTCCAGAAATGGAGCTCTCCC- 1504
               |||||||||||||||||||||||||||||||||||||||||||||||||
Db   481 TGTTATAAACCAGGTGCCCCTCCTGCCCACTCCAGAAATGGAGCTCTTCCCA 540

Qy   1505 AGCCTCCCATCTGACTACAGCCCAGGGTGGGGACTCAGCGGGGTATAAAAGCTAGTGTGACCC 1564
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   541 AGCCTCCCATCTGACTACAGCCCAGGGTGGGGACTCAGCGGGGTATAAAAGCTAGTGTGACCC 600

Qy   1565 TG 1566
               ||
Db   601 TG 602

XX
CC The present invention describes primer sets for synthesising 5602
CC full-length cDNAs defined in the specification. Where a primer set
CC comprises: (a) an oligo-dT primer and an oligonucleotide complementary
CC to the complementary strand of a polynucleotide which comprises one of
CC the 5602 nucleotide sequences defined in the specification, where the
CC oligonucleotide comprises at least 15 nucleotides; or (b) a combination
CC of an oligonucleotide comprising a sequence complementary to the
CC complementary strand of a polynucleotide which comprises a 5'-end
CC sequence and an oligonucleotide comprising a sequence complementary to a
CC polynucleotide which comprises a 3'-end sequence, where the
CC oligonucleotide comprises at least 15 nucleotides and the combination of
CC the 5'-end sequence/3'-end sequence is selected from those defined in
CC the specification. The primer sets can be used in antisense therapy and
CC in gene therapy. The primers are useful for synthesising polynucleotides,
CC particularly full-length cDNAs. The primers are also useful for the
CC detection and/or diagnosis of the abnormality of the proteins encoded by
CC the full-length cDNAs. The primers allow obtaining of the full-length
CC cDNAs easily without any specialised methods. AAH03166 to AAH13628 and
CC AAH13633 to AAH18742 represent human amino acid sequences; AAB92446 to
CC AAB95893 represent human cDNA sequences; and AAH13629 to AAH13632
CC represent oligonucleotides, all of which are used in the exemplification
CC of the present invention.
XX
SQ Sequence 783 BP; 157 A; 245 C; 217 G; 155 T; 9 other;

Query Match
Best Local Similarity 100.0%; Score 521; DB 22; Length 783;
Matches 521; Conservative 0; Pred. No. 3.3e-9?; Mismatches 0; Indels 0; Gaps 0;

```
Qy  1046 GTGATCCTGGTGTTGATGGAAAGGGACCCCCAAGGGCGTTGACTTCACAGCTGTTCGTGAT 1105
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     1 GTGATCCTGGTGTTGATGGAAAGGGACCCCCAAGGGCGTTGACTTCACAGCTGTTCGTGAT   60

Qy  1106 CTGCTCGTGTCGGTGGAGGGGTAGAAGCCCTGCACAGCCCTCGCATATCTGGGCACTGACG 1165
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    61 CTGCTCGTGTCGGTGGAGGGGTAGAAGCCCTGCACAGCCCTCGCATATCTGGGCACTGACG  120

Qy  1166 GTGGCCCAGCCTGTTCTGTTCCACATCGCCATTGCTCAGAATACAGAGACGCCCCAGGCT 1225
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   121 GTGGCCCAGCCTGTTCTGTTCCACATCGCCATTGCTCAGAATACAGAGACGCCCCAGGCT  180

Qy  1226 GTGCTGAAGACAGCACCAGCAGCCGCCTCCAAGGGAAGTTCCACTTCCACACCGTGACCATC 1285
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   181 GTGCTGAAGACAGCAGCCAGCCGCCTCCAAGGGAAGTTCCACTTCCACACCGTGACCATC  240

Qy  1286 CAGATCAGGGACTACTCGGAGGAGACATGAAGGACTGTCAGGCATGCCAGGGCCCCTCAGAC 1345
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   241 CAGATCGAGGACTACTCGGAGGAGACATGAAGGACTGTCAGGGCATGCCAGGGCCCCTCAGAC  300

Qy  1346 TGACTGCTCAGCCACCAGGCCACCAACTGGGGCATGAACAGGACCTGCAGGTGGCTGGGACTGAG 1405
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   301 TGACTGCTCAGCCACCAGGCCACCAACTGGGGCATGAACAGGACCTGCAGGTGGCTGGGACTGAG  360

Qy  1406 TGTCCCCCAGGCCCAGGCCCAGGCCCAGGCCTGTGTTATAAAACAGGTCCCCC 1465
         |||||||||||||||||||||||||||||||||||||||||||||||||||
Db   361 TGTGCCCCAGGCCCAGGCGCCAGGGACTTTGCCTACCCCAGCGTGTGTTATAAAACAGGTCCCCC  420

Qy  1466 TCCTGACCTCTGCCCCACTCCAGGAANTGGAGCTCTTCCCAGCCTCCATCTGACTACAGC 1525
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   421 TCCTGACCTCTGCCCCACTCCAGGAANTGGAGCTCTTCCCAGCCTCCATCTGACTACAGC  480

Qy  1526 CAGGGTGGGGACTCAGCGGGGTATAAAGCTAGTGTGACCCTG 1566
         ||||||||||||||||||||||||||||||||||||||||||
Db   481 CAGGGTGGGGACTCAGCGGGGTATAAAGCTAGTGTGACCCTG  521
```

DT 26-JUN-2001 (first entry)
XX
DE Human cDNA sequence SEQ ID NO:15557.
XX
KW Human; primer; detection; diagnosis; antisense therapy; gene therapy; ss.
XX
OS Homo sapiens.
XX
XX
EP1074617-A2.
XX
PD 07-FEB-2001.
XX
PF 28-JUL-2000; 2000EP-0116126.
XX
PR 29-JUL-1999; 99JP-0248036.
PR 27-AUG-1999; 99JP-0300253.
PR 11-JAN-2000; 2000JP-0118776.
PR 02-MAY-2000; 2000JP-0183767.
PR 09-JUN-2000; 2000JP-0241899.
XX
DR WPI; 2001-318749/34.
XX
PI Ota T, Isogai T, Nishikawa T, Hayashi K, Saito K, Yamamoto J;
PI Ishii S, Sugiyama T, Wakamatsu A, Nagai K, Otsuki T;
XX
PA (HELI-) HELIX RES INST.
XX
PS Claim 8; SEQ ID 15557; 2537pp + CD ROM; English.
XX
PT Primer sets for synthesizing polynucleotides, particularly the 5602
PT full-length cDNAs defined in the specification, and for the detection
PT and/or diagnosis of the abnormality of the proteins encoded by the
PT full-length cDNAs -
XX
CC The present invention describes primer sets for synthesizing 5602
CC full-length cDNAs defined in the specification. Where a primer set
CC comprises: (a) an oligo-dT primer and an oligonucleotide complementary
CC to the complementary strand of a polynucleotide which comprises one of
CC the 5602 nucleotide sequences defined in the specification, where the
CC oligonucleotide comprises at least 15 nucleotides; or (b) a combination
CC of an oligonucleotide comprising a sequence complementary to the
CC complementary strand of a polynucleotide which comprises a 5'-end
CC sequence and an oligonucleotide comprising a sequence complementary to a
CC polynucleotide which comprises a 3'-end sequence, where the
CC oligonucleotide comprises at least 15 nucleotides and the combination of
CC the 5'-end sequence/3'-end sequence is selected from those defined in
CC the specification. The primer sets can be used in antisense therapy and
CC in gene therapy. The primers are useful for synthesising polynucleotides,
CC particularly full-length cDNAs. The primers are also useful for the
CC detection and/or diagnosis of the abnormality of the proteins encoded by
CC the full-length cDNAs. The primers allow obtaining of the full-length
CC cDNAs easily without any specialised methods. AAH03166 to AAH13628 and
CC AAH13633 to AAH18742 represent human amino acid sequences; AAB92446 to
CC AAB95893 represent human cDNA sequences; and AAH13629 to AAH13632
CC represent oligonucleotides, all of which are used in the exemplification
CC of the present invention.
XX
SQ Sequence 1440 BP; 420 A; 412 C; 310 G; 0 T; 0 other;

Query Match
Best Local Similarity 100.0%; Score 521; DB 22; Length 1440;
Matches 521; Conservative 0; Pred. No. 3.9e-97; Mismatches 0; Indels 0; Gaps 0;

```
Qy  1046 GTGATCCTGGTGTTGATGGAAAGGGACCCCCAAGGGCGTTGACTTCACAGCTGTTCGTGAT 1105
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     1 GTGATCCTGGTGTTGATGGAAAGGGACCCCCAAGGGCGTTGACTTCACAGCTGTTCGTGAT   60

Qy  1106 CTGCTCGTGTCGGTGGAGGGGTAGAAGCCCTGCACAGCCCTCGCATATCTGGGCACTGACG 1165
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    61 CTGCTCGTGTCGGTGGAGGGGTAGAAGCCCTGCACAGCCCTCGCATATCTGGGCACTGACG  120

Qy  1166 GTGGCCCAGCCTGTTCTGTTCCACATCGCCATTGCTCAGAATACAGAGACGCCCCAGGCT 1225
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   121 GTGGCCCAGCCTGTTCTGTTCCACATCGCCATTGCTCAGAATACAGAGACGCCCCAGGCT  180
```

RESULT 6
AAH16516
ID AAH16516 standard; cDNA; 1440 BP.
XX
AC AAH16516;
XX

```
Db   121 GTGGCCCAGGCTGTTCTGTTGTGTCCACATGCCCATTGCTCAGGATAACAGAGACGCCCAGGCT 180
Qy  1226 GTGCTGAAGACAGCCACAGCAGCGCCGCCTCCAAGGGAAGTTCCACTTCCACACCGTGACCATC 1285
Db   181 GTGCTGAAGACAGCCACAGCAGCGCCGCCTCCAAGGGAAGTTCCACTTCCACACCGTGACCATC 240
Qy  1286 CAGATCGGAGGACTACTCGGAGGACATGAAAGGACTGTCAGGCATGCCAGGGCCCCTCAGAC 1345
Db   241 CAGATCGGAGGACTACTCGGAGGACATGAAAGGACTGTCAGGCATGCCAGGGCCCCTCAGAC 300
Qy  1346 TGACTGCTTCAGCCAGCCAGACCAACTGGGGGCATGAACAGGACCTGCAGGTGGGCTGGACTGAG 1405
Db   301 TGACTGCTTCAGCCAGCCAGACCAACTGGGGGCATGAACAGGACCTGCAGGTGGGCTGGACTGAG 360
Qy  1406 TGTCCCCCAGGCCCAGGCAGACTTTGCCTACCCCCAGCGCTGTGTTGTTATAAAACCAGGTCCCC 1465
Db   361 TGTCCCCCAGGCCCAGGCAGACTTTGCCTACCCCCAGCGCTGTGTTGTTATAAAACCAGGTCCCC 420
Qy  1466 TCCTGACCTCTGCCCACTCCAGGAATGGGAGCTCTTCCCAGCCTCCAGCTCTGACTACAGC 1525
Db   421 TCCTGACCTCTGCCCACTCCAGGAATGGGAGCTCTTCCCAGCCTCCAGCTCTGACTACAGC 480
Qy  1526 CAGGGTGGGGACTCAGCGGGTATAAAAGCTAGTGTGACCCTG 1566
Db   481 CAGGGTGGGGACTCAGCGGGTATAAAAGCTAGTGTGACCCTG 521
```

```
RESULT 7
ABN83947
ID   ABN83947 standard; DNA; 11101 BP.
XX
AC   ABN83947;
XX
DT   02-OCT-2002  (first entry)
XX
DE   Human transporter protein encoding genomic DNA.
XX
KW   Human; transporter protein; zinc transporter; pharmacogenomic analysis;
KW   diagnosis; drug screening; gene therapy; kidney; testis; heart; placenta;
KW   small intestine; liver; chromosome 1; single nucleotide polymorphism;
KW   SNP; gene; ds.
XX
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FT   CDS             2284..8840
FT                   /tag= a
FT                   /product= "transporter protein"
FT                   /note= "contains 7 introns"
FT   exon            2284..2333
FT                   /tag= b
FT                   /number= 1
FT   intron          2334..2962
FT                   /tag= c
FT                   /number= 1
FT   exon            2963..3183
FT                   /tag= d
FT                   /number= 2
FT   intron          3184..3735
FT                   /tag= e
FT                   /number= 2
FT   exon            3736..3882
FT                   /tag= f
FT                   /number= 3
FT   intron          3883..4629
FT                   /tag= g
FT                   /number= 3
FT   exon            4630..4783
FT                   /tag= h
FT                   /number= 4
FT   intron          4784..5472
FT                   /tag= i
FT                   /number= 4
FT   exon            5473..5632
FT                   /tag= j
FT                   /number= 5
FT   intron          5633..6375
FT                   /tag= k
FT                   /number= 5
FT   exon            6376..6481
FT                   /tag= l
FT                   /number= 6
FT   intron          6482..8088
FT                   /tag= m
FT                   /number= 6
FT   exon            8089..8223
FT                   /tag= n
FT                   /number= 7
FT   intron          8224..8694
FT                   /tag= o
FT                   /number= 7
FT   exon            8695..8838
FT                   /tag= p
FT                   /number= 8
XX
PN   WO200224910-A2.
XX
PD   28-MAR-2002.
XX
PF   20-SEP-2001; 2001WO-US29218.
XX
PR   20-SEP-2000; 2000US-234160P.
PR   19-OCT-2000; 2000US-0691219.
XX
PA   (PEKE ) PE CORP NY.
XX
PI   Wei M, Ketchum KA,  Di Francesco V,  Beasley EM;
XX
DR   WPI; 2002-404954/43.
DR   P-PSDB; ABB83084.
XX
PT   Novel human transporter proteins, related to zinc transporter
PT   subfamily, useful as model for developing human therapeutic targets and
PT   serves as target for human therapeutics -
XX
PS   Claim 4(b); Fig 3; 75pp; English.
XX
CC   The invention relates to an isolated human transporter protein that is
CC   related to the zinc transporter subfamily. Polynucleotides and
CC   polypeptides of the invention are useful for treating a disease or
CC   condition mediated by human transporter protein. The proteins also
CC   provide a target for diagnosing a disease or predisposition to disease
CC   mediated by the peptide, and in pharmacogenomic analysis. The peptides
CC   are also useful for treating a disorders characterised by absence of,
CC   inappropriate or unwanted expression of the protein. The nucleic acids
CC   are also useful in drug screening assays and as a target for treatment
CC   by the compounds identified through drug screening. The invention also
CC   provides vectors for gene therapy in patients with aberrant expression
CC   of the gene encoding the transporter protein. The gene of the invention
CC   has been found to be expressed in humans in the kidney, testis, heart,
CC   placenta, small intestine and liver. The gene has been localised to
CC   human chromosome 1. The current sequence represents the human transporter
CC   protein encoding genomic DNA.
CC   Note: This sequence contains 6 single nucleotide polymorphisms (SNP's),
CC   but information given in the specification is insufficient to determine
CC   their locations within this sequence (see ABN83948-ABN83953 for specific
CC   regions).
XX
SQ   Sequence 11101 BP; 2403 A; 3041 C; 3200 G; 2448 T; 9 other;
```

```
Query Match             22.5%;  Score 363.6;  DB 24;  Length 11101;
Best Local Similarity   98.9%;  Pred. No. 9.9e-65;
Matches  366;  Conservative   0;  Mismatches   4;  Indels   0;  Gaps   0;

Qy  1197 CCATTGCTCAGAATACAGAGACGCCCAGGCTGTGCTGTCGAAGACAGCCACAGCAGCGCCGCTCCAAG 1256
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
```

```
Db   8689  CCCCAGCTCAGAATACAGACGCCCGTGTGCTGAAGACAGCAGCCAGCCGCCCTCCAAG  8748
Qy   1257  GGAAGTTCCACTTCCACACCGTGACCATCCAGATCAGGACTACTCGGAGACATGAAGG  1316
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   8749  GGAAGTTCCACTTCCACACCGTGACCATCCAGATCAGGACTACTCGGAGACATGAAGG  8808
Qy   1317  ACTGTCAGGACATGCCAGGGCCCCTCAGACTGACTGCTCAGCCAGGACCAACTGGGCCAT  1376
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   8809  ACTGTCAGGACATGCCAGGGCCCCTCAGACTGACTGCTCAGCCAGGACCAACTGGGCCAT  8868
Qy   1377  GAACAGGACCTGCAGGTGGCTGGACTGAGTGTCCCCAGGGACTTTTGCCTA  1436
            ||||||||||||||||||||||||||||||||||||||||||||||||||
Db   8869  GAACAGGACCTGCAGGTGGCTGGACTGAGTGTCCCCAGGGACTTTTGCCTA  8928
Qy   1437  CCCCAGCTGTGTTATAAAAACCAGGTCCCCCTCCTCTGCCCCACTCTCCAGGAATGGAG  1496
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   8929  CCCCAGCTGTGTTGTTAAACCAGGTCCCCTCCTTGACCTCTGCCCCACTCTCCAGGAATGGAG  8988
Qy   1497  CTCTTCCCAGCCTCCCATCTGACTACACGCCAGGGTGGGGACTCAGGCGGGGTATAAAGCTAG  1556
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   8989  CTCTTCCCAGCCTCCCATCTGACTACACGCCAGGGTGGGGACTCAGGCGGGGTATAAAGCTAG  9048
Qy   1557  TGTGACCCTG  1566
            ||||||||||
Db   9049  TGTGACCCTG  9058
```

Db   8689
Qy   1257
Db   8749
Qy   1317
Db   8809
Qy   1377
Db   8869
Qy   1437
Db   8929
Qy   1497
Db   8989
Qy   1557
Db   9049

RESULT 8

ID   ABK71914 standard; cDNA; 1318 BP.
XX
AC   ABK71914;
XX
DT   30-JUL-2002 (first entry)
XX
DE   Human cDNA encoding zinc transporter-like protein NOV2c.
XX
OS   Homo sapiens.
XX
OS
KW   rheumatoid arthritis; Grave's disease; wound healing.
KW   hepatitis; fertility; hypertension; arteriosclerosis; ischaemia;
KW   haematopoietic disorder; gastrointestinal disease; respiratory disorder;
KW   inflammation; autoimmune disorder; immune disorder; blood disorder;
KW   angiogenesis; asthma; X-linked severe combined immunodeficiency;
KW   reproductive disorder; tissue disorder; thrombocytopaenia; migraine;
KW   central nervous system disorder; depression; lung disorder;
KW   neurodegenerative disorder; Alzheimer's disease; acute brain injury;
KW   vascular disorder; infectious disease; anorexia; cancer; stroke;
KW   Human; ss; gene; NOVX; developmental disorder; endocrine disorder;
XX

CC   The invention relates to an isolated NOVX polypeptide selected from
CC   NOV1a, NOV1b, NOV1ac, NOV2a, NOV2b, NOV2c, NOV3a, NOV3b, NOV4a, NOV4b,
CC   NOV5a, NOV5b or NOV6-NOV9 polypeptides, their mature form or variant.
CC   Also included are a nucleic acid encoding a NOVX protein or variant;
CC   a vector comprising the nucleic acid; a cell comprising the vector;
CC   an anti-NOVX antibody; and identifying agents that modulate the
CC   expression or activity of NOVX. The nucleic acid, antibody and
CC   modulators are useful in the diagnosis, treatment or prevention of
CC   developmental disorders, endocrine disorders, vascular disorders,
CC   infectious disease, anorexia, cancer, neurodegenerative disorders (e.g.
CC   Alzheimer's disease, Parkinson's disease, Huntington's disease, multiple
CC   sclerosis and amyotrophic lateral sclerosis), acute brain injury (e.g.
CC   stroke, head injury and cerebral palsy), central nervous system disorders
CC   (e.g. depression, epilepsy and schizophrenia), lung disorders
CC   reproductive disorders, disorders affecting carbohydrate metabolism (e.g.
CC   galactosaemia and hereditary fructose intolerance), tissue disorders
CC   (e.g. Wiskott-Aldrich syndrome, thrombocytopaenia, night blindness and
CC   Pick's disease), disorders linked to abnormal angiogenesis, asthma,
CC   azoospermia, learning disabilities, facial dysmorphism, autoimmune
CC   encephalomyelitis, X-linked severe combined immunodeficiency, seizures,
CC   migraines, inflammation, autoimmune disorders, disorders affecting sleep,
CC   appetite, thermoregulation, pain, perception, hormone secretion and
CC   sexual behaviour, immune disorders, haematopoietic disorders or other
CC   disorders related to cell signal processing and metabolic pathway
CC   modulation, gastrointestinal diseases, respiratory disorders, blood
CC   disorders, hepatitis, trauma, regeneration, viral, bacterial or parasitic
CC   infections, hyper- or hypo-thyroidism, endometriosis, fertility,
CC   hypertension, arteriosclerosis, ischaemia, haemolytic anaemia, Werner
CC   syndrome, rheumatoid arthritis, Grave's disease, wound healing, X-linked
CC   mental retardation, psychotic and neurological disorders and neuronal
CC   degeneration. The present sequence encodes a NOVX protein.
XX

SQ   Sequence 1318 BP; 337 A; 329 C; 308 G; 344 T; 0 other;

```
Query Match         19.1%;   Score 309.6;  DB 24;   Length 1318;
Best Local Similarity  57.9%;  Pred. No. 6.3e-54;
Matches  574;  Conservative  0;  Mismatches 409;  Indels   9;  Gaps    1;

Qy   359  GAGCTGGCTGCCCAGAGCAACCATCACTGCCATGCTCAGAAGAGGTCCTCGACAGTCACTGT  418
           |||||||  | |||||||  ||||||||  ||||||||||| || |||||||||||
Db   180  GAGCTGGACTCAGGAGGCATGTACCACTGCCACAGTGCCTCCAAGCCCACAGAAAAGGG  239
Qy   419  GACCCCAAGAGAAGGGGAAGGCCCAGCGCCTGTATGTAGCCTTCGCCATCTGCCTGTTG  478
           | ||||  | |||||||  ||||||||  || |||  ||||  ||||||||
Db   240  GCGAATGAGTACGCCTATGCCAAAGTGGGAACCTGTTCTGCTTCACAATATGCTTCATT  299
Qy   479  TTCATGATCGGAGAAGTCGTTGGTGGGTGTACCTGGCACACAGCTTGGCTGTCATGACTGAC  538
           ||||||||| |||||||| |||| ||| ||||  ||||||| || |||||||
Db   300  TTCATGAATTGCAGAGGTCGTGGGTGGGGCAACATGGCTGGGGAGTCGTTGCTGTCTGCACAGAT  359
Qy   539  GCAGCACACCTGCTCACTGACATTTGCCAGCATGCTCATCACGCCTCTTCTCCCCTCTTGATG  598
           ||||||  || |||||  |||||| ||||||  ||  |||||||||  ||||||||||||
Db   360  GCTGCCCACCTCTTTAATTGACCTGACCAGTCTCTGCCTCAGGTCTCTTCTCCCCGTGGTTG  419
Qy   599  TCCTCCCCGGCCAGCCCACCAAGACCCATGAACTTTGGGCTGGCAGAGAGCTGAGACTCGGGA  658
           | || |||||||  ||||  ||||||||||||  ||||||||||  |||||||||||
Db   420  TCATCGAAGCCTCCCTCTAAGCGGCTGCACATTTGGGATGGCACCGAGAGCAGAGATCCTTGGT  479
```

```
Qy   659 GCCCTGGCTGCTCTCGTGCTACTGTGTGTCATCTCGGGTGTCGTTGACCGGGGGTACTGGTGTGTGTGTACTGGCTGTG  718
         |||||| | |  ||  ||| |  ||||| |  ||  ||| ||| ||  || ||||||  | |     |||  ||||| |
Db   480 GCCCTGCGTCTCCATCCTGTGCATCTGGGTGTTGGTGGCTGGTTGGTCGGCTGGCTGGTCTAGTTGTTGTACCTGGCATGT  539

Qy   719 GACGCGGCTGATCTCTGGGGACTATGAAATTGACGGGGGGGGACCATGCTGATCACGTCGGGC  778
         |||| |  |||||  | | | || | | | || |||||| || |  | | |  ||| |||||
Db   540 GAGCGGCTGCTGCTGTATCCTGATTACCAGATCCAGGCGACTGTGATGATCATCGGTTCCAGC  599

Qy   779 TCGCTGCTGGGCTGTGAACATCATAAATGGGGTTGACCCTTCACCAGTGTCGGCCATGGGCAC  838
         |||| || ||   |  | | | |   |||||  ||  ||  | | ||||   |||||||| |
Db   600 TCGGCAGTGGCGGCGGCCAACATTGTTACTAACTGTGGTTTTGCACCAGAGATGCCTTGGGCAC  659

Qy   839 AGCCACGGGCACCCAACCACGCAGGAGGAGGAACCCCAGCGTCCGAGCCTGCCTTCATCCAT  898
         |||| || ||||  | |  |  |||  | |  |  ||| |||| |||| ||  |||| ||||
Db   660 CAAGGAAGTACAAGCCCAATGCCACGGTCAGAGCTGCTTGTTTTGTCAT  710

Qy   899 GTGATCGGCGACTTTATGCAGGAGCATGGGTGTGTCCTAGTGGCAGCCTATATTTTTATACTTC  958
         |||| || | || | || ||| | || | |||| |  ||| ||| | |  | | | ||| ||
Db   711 GCCCTTGGAGATCTATTTCAGAGTATCAGTGTGTGCTAATTAGTGCACTTATTATTCTACTTT  770

Qy   959 AAGCCACAGAATACAAAGTATGTAGACCCCATCCTTCGTGTCTTCTCCATCCTGGGTCCTG  1018
         |||||||| | | || || |   | |||  || |  | ||| |||||||||  ||| |||
Db   771 AAGCCAGAGTATATAAAAATTAGGGCACCCAAATCTGCACATTCATCTTTTCCATCCTGGGTCTTG  830

Qy  1019 GGGACAAACCTTGACCATCCTGAGAGACGATGTGATCCTCGGTGTTGTGATGGAAGGACACCCCAAG  1078
         |||| ||  || || || |  ||||||||| |||  |||| ||||| ||| |  | ||||||||||
Db   831 GCCAGACCATCACTACTATCTTAAAGGACTTCTCCATCTTACTCATGGAAGGGTGTGCCAAAG  890

Qy  1079 GGCGGTTGACTTCACAGCGTTCGTGATCGCTGCTGTTCGTGGGGGGGTAGAAGCCCTG  1138
         ||| ||  | ||  | ||  |  |||  | ||| | | ||| || |  ||||  ||||
Db   891 AGCCTGAATTACAGTGGTGTGAAAGAGCTTATTTAGCAGTCGCACCTGTGTGTGTGTG  950

Qy  1139 CACAGCCTGCTGCATATCTGGGCACTGACGGTGGCCCAGCGTCGTGTTCGTGTCCACATCGCC  1198
         |||||| ||| || | | | | || || ||| ||| |  | |  ||||| || || |||||
Db   951 CACAGCCTGCACATCTGGTCTCTAACAAATGAATCAAGTAATTCTTCAGCTCATGTTGCT  1010

Qy  1199 ATTGCTCAGAATACAACAGCGCCCAGGGTTGCTGTTCGTGAAGACAGCAGAGCCGGCCTCCCAAGGG  1258
         |||| | | ||| | | ||| | ||  ||| | || ||||||| | | | ||||  |  |||||
Db  1011 ACAGCAGCCAGCCGGGACAGCCAAGTGGTTCGGAGAGAAATTGCTAAAGCCCTTAGCAAA  1070

Qy  1259 AAGTTCCACTTCACACGTGACCATCCAGATCGAGGACTACTCGGGGACATGAAGGAC  1318
         |||| |  | ||  | || | ||| |||  | || |||| | | | |    || |||
Db  1071 AGCTTTACGACGCACTCACTCACCATTCAGATAGGAATCTCCAGTTGACCAGGACCCGAC  1130

Qy  1319 TGTCAGGCATGCCAGGGCCCCTCAGACTGACT  1350
         | |  |  ||||| | |  |||| |||||||
Db  1131 TGCCTTTCTGTGAAGACCCCTGTGACTAGCT  1162
```

RESULT 9
ABK71912

ID   ABK71912 standard; cDNA; 1431 BP.
XX
AC   ABK71912;
XX
DT   30-JUL-2002 (first entry)
XX
DE   Human cDNA encoding zinc transporter-like protein NOV2a.
XX
KW   Human; ss; gene; NOVX; developmental disorder; endocrine disorder;
KW   vascular disorder; infectious disease; anorexia; cancer; stroke;
KW   neurodegenerative disorder; Alzheimer's disease; acute brain injury;
KW   central nervous system disorder; depression; lung disorder;
KW   reproductive disorder; tissue disorder; thrombocytopaenia; migraine;
KW   angiogenesis; asthma; X-linked severe combined immunodeficiency;
KW   inflammation; autoimmune disorder; immune disorder; blood disorder;
KW   haematopoietic disorder; gastrointestinal disease; respiratory disorder;
KW   hepatitis; fertility; hypertension; arteriosclerosis; ischaemia;
KW   rheumatoid arthritis; Grave's disease; wound healing.
XX
OS   Homo sapiens.
XX
PN   WO2002024733-A2.
XX
PD   28-MAR-2002.
XX
PF   17-SEP-2001; 2001WO-US29115.
XX
PR   15-SEP-2000; 2000US-232675P.
PR   15-SEP-2000; 2000US-232676P.
PR   15-SEP-2000; 2000US-232679P.
PR   18-SEP-2000; 2000US-233382P.
PR   18-SEP-2000; 2000US-233402P.
PR   19-SEP-2000; 2000US-233521P.
PR   19-SEP-2000; 2000US-233522P.
PR   19-SEP-2000; 2000US-233801P.
PR   20-SEP-2000; 2000US-233960P.
PR   06-OCT-2000; 2000US-238398P.
PR   06-OCT-2000; 2000US-238399P.
PR   13-OCT-2000; 2000US-240284P.
PR   13-OCT-2000; 2000US-240498P.
PR   11-JAN-2001; 2001US-260973P.
PR   26-JAN-2001; 2001US-264274P.
PR   09-MAR-2001; 2001US-274862P.
XX
PA   (CURA-) CURAGEN CORP.
XX
PI   Mishra VS, Syptek KA, Taupier RJ, Vernet CAM, Colman SD, Gorman L;
PI   Tcherney VT, Malyankar UM, Shenoy S, Tcherney VT, Padigaru M;
PI   Patturajan M, Burgess CE, Smithson G, Millet I, Peyman JA;
PI   Stone D, Gunther E, Ellerman K;
XX
WPI; 2002-383182/41.
DR   P-PSDB; ABG660224.
XX
PT   New cytoplasmic, nuclear, membrane bound and secreted NOVX
PT   polypeptides, useful for treating cancers and tumours, lung disorders,
PT   haematopoietic disorders, autoimmune diseases and immune disorders -
XX
PS   Claim 9; Page 19-20; 210pp; English.
XX
CC   The invention relates to an isolated NOVX polypeptide selected from
CC   NOV1a, NOV1b, NOV1ac, NOV2a, NOV2b, NOV2c, NOV3a, NOV3b, NOV4a, NOV4b,
CC   NOV5a, NOV5b or NOV6-NOV9 polypeptides, their mature form or variant.
CC   Also included are a nucleic acid encoding a NOVX protein or variant;
CC   a vector comprising the nucleic acid; a cell comprising the vector;
CC   an anti-NOVX antibody; and identifying agents that modulate the
CC   expression or activity of NOVX. NOVX, the nucleic acid, antibody and
CC   modulators are useful in the diagnosis, treatment or prevention of
CC   developmental disorders, endocrine disorders, vascular disorders,
CC   infectious disease, anorexia, cancer, neurodegenerative disorders (e.g.
CC   Alzheimer's disease, Parkinson's disease, Huntington's disease, multiple
CC   sclerosis and amyotropic lateral sclerosis), acute brain injury (e.g.
CC   stroke, head injury and cerebral palsy), central nervous system disorders
CC   (e.g. depression, epilepsy and schizophrenia), lung disorders,
CC   reproductive disorders, disorders affecting carbohydrate metabolism (e.g.
CC   galactosaemia and hereditary fructose intolerance), tissue disorders
CC   (e.g. Wiskott-aldrich syndrome, thrombocytopaenia, night blindness and
CC   Pick's disease), disorders linked to abnormal angiogenesis, asthma,
CC   azoospermia, learning disabilities, facial dysmorphism, autoimmune
CC   encephalomyelitis, X-linked severe combined immunodeficiency, seizures,
CC   migraines, inflammation, autoimmune disorders, disorders affecting sleep,
CC   appetite, thermoregulation, pain, perception, hormone secretion and
CC   sexual behaviour, immune disorders, haematopoietic disorders or other
CC   disorders related to cell signal processing and metabolic pathway
CC   modulation, gastrointestinal diseases, respiratory disorders, blood
CC   disorders, hepatitis, trauma, regeneration, viral, bacterial or parasitic
CC   infections, hyper- or hypo-thyroidism, endometriosis, fertility,
CC   hypertension, arteriosclerosis, ischaemia, haemolytic anaemia, Werner
CC   syndrome, rheumatoid arthritis, Grave's disease, wound healing, X-linked
CC   mental retardation, psychotic and neurological disorders and neuronal
CC   degeneration. The present sequence encodes a NOVX protein.
XX
SQ   Sequence 1431 BP; 373 A; 345 C; 330 G; 383 T; 0 other;

Query Match   19.1%; Score 309.6; DB 24; Length 1431;
Best Local Similarity   57.9%;
                        Pred. No. 6.5e-54;

RESULT 10
ABK88011

Matches 574; Conservative 0; Mismatches 409; Indels 9; Gaps 1;

```
Qy    359 GAGCTGGCCTGCCCAGAGCAACCATCACTGCCATGCTCTCAGAAGGTGCTTGACAGTCACTGT 418
          ||||||| |      |   |   || ||||||
Db    421 GAGCTGGAGTCAGGAGGCATTGTACCACTGCCTCCACAGTGGGCTCCAAGCCACAGAAAAGGG 480

Qy    419 GACCCCAAGAGAGGGGAAGGCCCAGCGCCCACTGTATGTAGCCTGTTCCCATCTGCCTGTTG 478
          | |    || ||| |   ||  |  ||     |   || |  || | |  ||  ||| |||| 
Db    481 GCGAATGAGTACGGCCTATGCCAAGTGGAAAACTCTGTTCTGCTTCACCAATATGCTTCATT 540

Qy    479 TTCATGATCGGAGAGTACGTTTGGTGGGTACCTGGCACACAGCTTGGCTGTCATGACTGAC 538
          |||||||| | |||| | || | ||||| |   | | ||  |   ||||||| ||||| ||
Db    541 TTCATGATTGCGAGAGTCGTGGGTGGGGCACATTGTGGGAGTCTTGGCTGTGTCACAGAT 600

Qy    539 GCAGGACACCTGCTCACTGACTTTGCCCAGCAATGCTCATCAGCCTTCTCCCCTCTGGATG 598
          ||  |     | || |  |  || || |||   ||   || |   |  |   | ||| | |
Db    601 GCTGCCACTTCTTAATTGACTGACCAGTTTCCTGCTCAGTCTTCTTCTCTCCCTGTTGGTTG 660

Qy    599 TCCTCCCGGCCAGCCACCAAGACCATGAACTTGCCTGTGACAGTCTGGGA 658
          | ||| |  |||| |  || || ||| | |  |  || | ||| ||||||
Db    661 TCATCGAAGCTCCCTCTAAGCGGCTGACAATTGGATGGCACCAGAGATCCTTGGT 720

Qy    659 GCCCTGGTCTCTGTACTGTCCATCTGGGTCTTGTCTGACGGGGTTACTGTTGTTACCTGGGTGTG 718
          || ||| |   |||  | || |||| |||  || |  ||  | |  |  |   | ||||| |||||
Db    721 GCCCTGCTCTCCATCCTGTGTGACTCGGGTGTGGTTGCTCGCTGCTGGCTGCTCTACCTGGGCATGT 780

Qy    719 GAGCGGCTGATCTCTGGGGACTATGAAAATTGACGGGGGAGACAGTGAGATCTTGGGA 778
          |||| | || ||||| |  |  |  |  |||  |  | |||||||||| |  ||| |||
Db    781 GAGCGGCCTGCGTGTATCCTGATTACCAGACTGTGATGATCATCGTTGTGGAGTCAGC 840

Qy    779 TGCGCTGTTGGGTGTGAACATCATAATGTGGGTTGACCCCTTCACCAGTGTGACTGGGCGAC 838
          || ||||| || | || ||||  | | ||| ||||||| || | ||| || || | ||| |
Db    841 TGCGCAGTGGGCGGCCCAACATTGTACTAACTGTTGGTTTTGACACTGAGAGATCCTTGGCAC 900

Qy    839 AGCCACGGCACCACCAAACCACCAGGAGGAGAACCCCAGCGTCGGACCTGCCTGCATGT 898
          || || |  | || || |  |  ||||| | |  |||  |  ||| | | | ||| ||
Db    901 CAAGGAAGTACAAGCCAAATGCCAGGTGTCAGAGCCTGCAGACCTGCTCTTCTCATCCAT 951

Qy    899 GTGATCGGGCGGACTTTATGCAGAGACAGCATGGGGTGTCCTAGTGGCCAGCCTATATTTTATACTTC 958
          ||| | | | ||||| ||| |||| ||| | |  | || || |  |  ||  || || |  | ||
Db    952 GCCCTTGGAGAGATCTATTTCAGTGTTGCTAATTAGTGCACTTATTATTATCTACTTT 1011

Qy    959 AAGCCAGAAATACAAGTATGTGAGAGACCCCATCTGCACCTTCGTCTTCTCCCATCCTGGTCCTG 1018
          || || ||  |  | | || ||||| || ||| ||| || ||| || || |  | ||| |  |||
Db   1012 AAGCCAGAGTATAAAATAGCCGACCCAAATCTGCACAATCGCACATCTTTTCCATCCTGGTCTTG 1071

Qy   1019 GGGACAAACCTTGACCCATCCTGAGAGATGTGATCCTGGTGTTGTTGATGGAAGGACACCCCAAG 1078
          |||| | |  ||| | ||||  | | || |  |  |||| | || |||||| || | | || ||
Db   1072 GCCAGCACCATCACTATCTTAAAGGACTTCTCCATCATGGAAGGTTGTGCCAAAG 1131

Qy   1079 GGGCGTTGACTTCACACGTTTTCGTCGTCTCTGCTCTCGTCGTGGGAGGGGGGTAGAAGCCCTG 1138
          ||| | || |  |  || ||| | | | | |  |  |  | | || ||| || | | | | |||
Db   1132 AGCCTGAATTACAGTGGTGTGAAAGAGCTTATTTTTACAGTCGGACGGGTGTCTGTCTG 1191

Qy   1139 CACAGCCTGCATATCTGGGCACTGACGGGGCCGCCAGCGTTGTCGTTCCACATCGGCC 1198
          ||| ||| || |  | |  | | | | || |   ||| | | ||  | || |   |||
Db   1192 CACACGCCTGCACATCTGGGTCTTCTAACGAATGAATCAAGTATTCTCTCAGCTCAGTGTTGCT 1251

Qy   1199 ATTGCTCAGAATAACAGAGACGCCCAGCGTGTCGTGCTGAAGACAGCAGCCGCTCCAAGGG 1258
          | |  | |   |  |  | |  | || | || | || | |  | | || |  | | |   |
Db   1252 ACAGCACAGCTGGGACAGCCCAAGTGGTTCGGAGGACTGGAAAATTGCTCTAAAAGCCCTTAGCAAA 1311

Qy   1259 AAGTTCCACTTCCACACCGTGACCCATCCAGATCGACGAGGACTACTCCGGAGGACATGAAGGAC 1318
          ||| |  |   |     | | || | ||| || |  ||| | |  || ||||||||| |||||||
Db   1312 AGCTTTACGATGCACTCACTCACCAATTCAGATGGAATCCTGACGTTGACCAGGACCCCGAC 1371

Qy   1319 TGTCAGGCATGCCAGGGCCCCTCAGACTGACT 1350
          || ||  ||||| ||||||| | ||| |||||
Db   1372 TGCCTTTTCTGTGAAGACCCTGTGACTAGCT 1403
```

ID   ABK88011 standard; cDNA; 2165 BP.
XX
AC   ABK88011;
XX
DT   07-OCT-2002 (first entry)
XX
DE   cDNA encoding human 84233 metal transporter protein.
XX
KW   84233; metal transporter; human; ss; gene; infection;
KW   haematopoeilic disorder; blood clotting disorder; cancer;
KW   autoimmune disorder; leukaemia; immunological disorder;
KW   cardiovascular disorder; neurological disorder; cellular proliferation;
KW   red blood cell disorder; viral disease; neurological disorder.
XX
OS   Homo sapiens.
XX
FH   Key   Location/Qualifiers
FT
FT   CDS   526..1488
FT   /*tag= a
FT   /product= "84233 protein"
FT   /note= "This sequence is specifically
FT   claimed in claim 1 of the specification"
XX
PN   WO020240656-A2.
XX
PD   23-MAY-2002.
XX
PF   14-NOV-2001; 2001WO-US45291.
XX
DR   WPI; 2002-508325/54.
DR   P-PSDB; AAU99907.
XX
PI   Meyers RE, Curtis RAJ, Glucksmann MA;
XX
PA   (MILL-) MILLENNIUM PHARM INC.
XX
PS   Claim 1; Page 236-237; 298pp; English.
XX
CC   This invention relates to the DNA and protein sequences of novel
CC   isolated 47476, 67210, 49875, 46842, 33201, 83378, 84233, 64708, 85041
CC   or 84234 proteins. The method of the invention is useful for treating a
CC   disorder characterised by aberrant activity of 47476, 67210, 49875,
CC   46842, 33201, 83378, 84233, 64708, 85041 or 84234-expressing cell in a
CC   subject. The protein molecules can act as novel diagnostic targets and
CC   therapeutic agents for controlling aberrant or deficient signal
CC   transduction resulting in e.g., haematopoeitic disorders, including
CC   blood clotting disorders, autoimmune disorders, or disorders related to
CC   an inability to clear infections (e.g., viral or bacterial infections),
CC   as well as disorders related to abnormal cellular proliferation or
CC   differentiation, e.g., leukaemia. They may also be used to control
CC   disorders of metabolic imbalance (e.g., disorders of lippolysaccharide
CC   biosynthesis or glycogen synthesis) Immunological disorders,
CC   cardiovascular disorders, neurological disorders, or cellular
CC   proliferation and/or differentiation disorders, e.g., cancer, cell
CC   motility and adhesion disorders differentiative disorders, red blood
CC   cell disorders, viral diseases, neurological disorders (e.g., brain
CC   disorders), pain or metabolic disorders, liver disorders, kidney
CC   disorders, disorders of the small intestine, disorders of metal ion
CC   imbalance, protein trafficking disorders and disorders associated with
CC   bone metabolism. The sequences of the invention are also useful for
CC   screening assays, predictive medicine (e.g., diagnostic assays,

CC prognostic assays, monitoring clinical trials, and pharmacogenetics);
CC and methods of treatment (e.g. therapeutic and prophylactic). The
CC present sequence represents the cDNA encoding the human 84233
CC metal transporter protein of the invention.
XX
SQ Sequence 2165 BP; 664 A; 470 C; 454 G; 577 T; 0 other;

Query Match
Best Local Similarity 19.1%; Score 309.6; DB 24; Length 2165;
Matches 574; Conservative 57.9%; Pred. No. 7.2e-54; Indels 9; Gaps 1;
Mismatches 409; 0 other;

```
Qy  359 GAGCTGGCTGCCCAGAGCAACCATCACTGCCATGCTCAGAAGGGTCCTGACAGTCACTGT 418
           |||||||  ||  |  || |       |||       |||    ||     |  |||
Db  508 GAGCTGGAGTCAGGAGGCATGTACCACTGCCATCAGTGCCTCCAAGCCCACAGAAAAGGGG 567

Qy  419 GACCCCAAGAGAGGAGAAGGCCCCAGCGTGTATGTAGCCTCCCATCTGCCTGTTG 478
           || | |     |    ||  |  |   || |     |  | || |  | |||
Db  568 GCGAATGAGTAGCGCCTATGCCAAGTGGAAACTCTGTTCTGCTTCAGCAATATGCTTCATT 627

Qy  479 TTCATGATCGGAGAGAGTCGTTGGTGGTGGGTACCTGGCACACAGCTTGGGTGTCATGACTGAC 538
           |||||||  | |||||| |||| |||| ||| |   |  ||| | || |   | | | ||
Db  628 TTCATGATTTGCAGAGGTCGTGGGTGGGCACACAGTGTGGGAGTGTTGCCTGTTGTCACAGAT 687

Qy  539 GCAGCACAACCTGCTCACTGACTTTGCCAGCATGCTCATCAGCCTCTTCTCCCTCTGGATG 598
          ||||||||  ||  |     |   |||   ||  |  |  |  | ||  |||    | |||
Db  688 GCTGCCCACCTCTTAATTGACCTGACCGCAGTTTCCTGCTTCCAGTCTCTCCCTGTGGTTG 747

Qy  599 TCCTCCCGGCCAGCCACCACAGAACTTTGGGCTGCAGAGACTGAGATCTTGGGA 658
           ||    |||  | || | |  | || || |  |||  |  ||||| |||| | |
Db  748 TCATCGAAGCCTCCCTCTAAGCGGCTTGAACATTTGGATGGCACCGGAGATCCTTGGT 807

Qy  659 GCCCTGGTCTCTGTACTGTCCATCTGGGTGTCTGACGGGGGGTTACCTGGGCTGTG 718
           ||| |||||| ||  | |  || |||| | || ||  |||||  |  |||| ||||
Db  808 GCCCTGCTCTCCAGCCTCTGTGCATCTGGGTGCTGACTCTGGTGCTGTGTGGCCATGT 867

Qy  719 GAGCGGCTGATCTCTGGGACTATGAAATTGACGGGGGGACCATGCTGATCACGTCGGGC 778
          |||| ||||||| |   |  ||  |  ||  | |||    ||||| ||||  | |||||
Db  868 GAGCGGCTGCTGTATCCTGATTACCAGATCGTGATGATCATCCGTTTCCCAGC 927

Qy  779 TGCGCGTGTGGCTGTGGTGAACATCATAATGGGGGTTGACCCTTCACCAGGTCTGGCGGCAC 838
          |||| | ||||| |  ||||  ||||| | ||| ||||| ||| ||||||| | ||| |||
Db  928 TGCGCAGTGGCGGCCAACATGTGTACTAACTGTGGTTTTGCACCAGAGAATGCCTTGGGCCAC 987

Qy  839 AGCCACGGCACCACCAACCAGCAGGAGGAGAAACCCAGCGTCCGAGCTGCCTTCATCCAT 898
          |||||  ||| ||||  ||||||  | ||  |  || || |||  | |||| ||||||||
Db  988 AACCAGAGTGTATAAAATAGCCGACACCCAATCTCACCATTTTCCATCCTGGTTTG 1038

Qy  899 GTGATCGCGGACTTTATGCAGAGCATGGGTGTGTCTAGTGGTGGCAGCCTATATATTTATTACTTC 958
           ||  |  | |||    |||||||||||| ||| || |||| | ||| |||| | |  | | ||
Db 1039 GCCCTGGAGATCTATTTCAGAGTATCAGTGCTGCTATTTAGTGTGCACTTATTATTATCTACTTT 1098

Qy  959 AAGCCAGAAATACAAGTATGTTGAGACCCCATCTTCGACCCCCTTTCGGTTCCTG 1018
          |||||||  || |  | | |  || | || | | | | | ||| ||  | ||||
Db 1099 AAGCCAGAGTATAAAATAGCCGACCCAAATCTGCACCATTTCCCATCCTGGTTTG 1158

Qy 1019 GGGACAACCTTGACCATCCTGAGAGATGTGATCCTGGTGTTGCATGGAAGGACCCCCAAG 1078
          ||||| ||  ||    | |  | ||| |  | ||| || | | |||||| | || | |||
Db 1159 GCCACCATCACTATCTTAAAGAGCTTCTCCACATCTTATTAGTGCATGGTTGCCAAAG 1218

Qy 1079 GGCGTTGACTTCACAGCTGTTGTCGTGATCTGCGTGTCGGAGGGGTAGAAGCCCTG 1138
           ||| |||||    ||| || || |  || ||| |||  |  | |||| ||| ||||
Db 1219 AGCCTGAATTACAGTGGTGTGAAAGAGCTTATTTTAGCAGTCGAAAGAGCTGTGTG 1278

Qy 1139 CACAGCCTGCAATATCTCGGGGACTGACGGGTGGCCCAGCCTGTTCTGTTGTCTGCCACATCGCC 1198
          || |||||||   ||  | ||| ||  ||||  |||||||| || |  | | |  ||||| ||||
Db 1279 CACAGCCTGCACTTCTGGTCTGTTAACAATGAACTCAGTAATTCTCTTCAGCTTATGGT 1338

Qy 1199 ATTGGCTCAGAATACAGACGCCCAAGTGTCTGCTGAAGACAGCAGCCGCCTCCAAGGG 1258
          || | |   ||  | ||  |||| || ||| |||| ||| | ||| | |||| ||||
Db 1339 ACAGCAGCAGCCGCGGGACACCCAAGTGGTTCGGAGAGAAATTGCTAAAGCCCTTAGCAAA 1398
```

```
Db 1399 AGCTTTACGGATGCACTCACTCACCATTCACCATTCAGATGGAATCTCCAGTTGACCAGGAGACCCCGAC 1458
           ||  |   | || ||  |    ||| ||  | |  |||||  ||||  |||||||||||||
Qy 1319 TGTCAGGCATGCCAGGCGCCCCTCAGCACTGACT 1350
          ||  |||||  ||| |||||| |||||  ||||||
Db 1459 TGCCTTTTCTGTGAAGACCCTGTGACTAGCT 1490
```

RESULT 11
ABV23919
ID ABV23919 standard; cDNA; 2297 BP.
XX
AC ABV23919;
XX
DT 16-SEP-2002 (first entry)
XX
DE Human prostate expression marker cDNA 23910.
XX
KW Human; prostate cancer; cytostatic; carcinogen; pharmacodynamic marker;
KW pharmacogenomic marker; gene; ss.
XX
OS Homo sapiens.
XX
PN WO200160860-A2.
XX
PD 23-AUG-2001.
XX
PI 20-FEB-2001; 2001WO-US05171.
XX
PF
PR 17-FEB-2000; 2000US-183319P.
PR 16-MAR-2000; 2000US-189862P.
PR 25-MAY-2000; 2000US-207454P.
PR 09-JUN-2000; 2000US-211314P.
PR 18-JUL-2000; 2000US-219007P.
PR 13-DEC-2000; 2000US-255281P.
XX
PA (MILL-) MILLENNIUM PREDICTIVE MEDICINE INC.
XX
PI Schlegel R, Endege WO, Monahan JE;
XX
DR WPI; 2001-662795/76.
XX
PT Novel isolated nucleic acid molecule associated with cancerous state of
PT prostate cells and correlating with presence of prostate cancer, useful
PT for detecting presence of prostate cancer, stage of prostate cancer -
XX
PS Claim 1; Page 4416-4417; 11750pp; English.
XX
CC The invention relates to an isolated nucleic acid molecule (I) comprising
CC a nucleotide sequence given in Tables 1-9 (ABV00010-ABV62213) of the
CC specification or its complement. (I) is useful for:
CC (a) assessing whether a patient is afflicted with prostate cancer;
CC (b) monitoring the progression of prostate cancer in a patient;
CC (c) assessing the efficacy of a test compound to inhibit prostate
CC cancer in a patient;
CC (d) assessing the efficacy of a therapy for inhibiting prostate cancer
CC in a patient;
CC (e) selecting a composition for inhibiting prostate cancer in a patient;
CC (f) assessing the prostate cell carcinogenic potential of a compound;
CC (g) determining whether prostate cancer has metastasized in a patient;
CC (h) assessing the aggressiveness or indolence of prostate cancer in a
CC patient;
CC (I) is also useful as a pharmacodynamic or pharmacogenomic marker.
XX
SQ Sequence 2297 BP; 506 A; 491 C; 601 T; 7 other;

Query Match
Best Local Similarity 19.1%; Score 309.6; DB 23; Length 2297;
Matches 574; Conservative 57.9%; Pred. No. 7.3e-54; Indels 9; Gaps 1;
Mismatches 409; 0 other;

```
Qy  359 GAGCTGGCTGCTGCCCAGAGCAACCATCACTGCCATGCTCAGAAGGGTCCTGACAGTCACTGT 418
          |||||||  ||  ||    |   |  ||      ||||      ||   |    ||||| |||
Db  498 GAGCTGGAGTCAGGAGGCATGTGTACCAACACCAGTGCCTCCACAGTGCCTCAGCAGTCACTGT 557
```

```
Qy   419  GACCCCAAGAGAAGGGAAGGCCCAGCGTGATGTAGGCCTCTGCCATCTGCCTGTTG 478
          | |  ||  ||  || |   ||| || ||  || || || || || |||| |
Db   558  GCGAATGAGTACGCCCTATGCCAAGTGGAAAACTCTGTTGCTTCAGCAAATATGCTTCATT 617

Qy   479  TTCATGATTGCGAGAAGTCGTTGGTTTGGTGGTCACACAACTTGGCTTGCTGCGTCATGACTGAC 538
          |||| ||| || ||  |   |||| ||||| || || ||| ||| | || ||||  || |
Db   618  TTCATGATTGCAGAGGTCGTGGGTGGGCACACAACTTGGCTTGCTGCGTCATGACTTGTCACAGAT 677

Qy   539  GCAGACAACCTGCTCACTGACTTTGCCAGCAATGCTCATCAGCCTCTTCTCCCTCGGATG 598
          |||| ||  ||| || |    |||| ||  || || ||||| |  || || | |  ||
Db   678  GCTGCCAACTTCTTAATTGACCTGACCAGTTTTCCTTCTCAGTCTCTTCTCTCTCGGTTG 737

Qy   599  TCCTCCCCGGCCAGCCACCAAGACCCATGAAACTTTGGCTTGCTGCAGAGAAGCTGAGGATCTTGGGA 658
          |  | |  || ||| ||||||| ||  |  ||| |||||||||| || |||   ||| ||||
Db   738  TCATCGAAGCCTCCCTCTTAAGCGGCGTGACATTTGGATGGCACCGAGCCAGAGAATCCTTGGT 797

Qy   659  GCCCTGGTCTCTGTACTTGTCCATCTCTGGGTCGTCTGACGGGGTACTTGGTTGGCTGTG 718
          ||| || |  | |  || ||| | |   ||| |    ||||| ||  || ||  || |
Db   798  GCCCTGCTCTTCCTCCATCCTGTCTGGGTGGTTGGCTGGCTACTGGTGTAGCTGGCCATGT 857

Qy   719  GAGCGGCGCTGATCTCTGGGGACTATGAAAATTGACGGGGGGACCATCTGATCACGTCGGGC 778
          || ||| |||| | | |  ||||      |   | | ||    |  | || | || |
Db   858  GAGCCGCCTGCTGTATCCTGATTACCAGATCAGCGGTGCTGCTGATCATCTGTTTCCAGC 917

Qy   779  TGCGCTGTGGCCTGTGAACATCATAATAATGGGGTTGACCCTTCACCAGTCTGGGCCATGGGGCAC 838
          ||  | ||| ||  || || || |    || ||   |||   | |||||||| |  || |
Db   918  TGCCCAGTGGCGGCGGCCAACATTGTACTAAACGTGGTTTTTGCACCAGAGATGGCCTTGGGCAC 977

Qy   839  AGCCACGGCGCCACCAACCAGCACCAGGAGGAGAACCCCGAGGGTTACTTGGTTGGCTGTGTG 898
          || |  | | |    ||||  || ||||| |  | |  ||||| || |||  ||  ||
Db   978  AATCA------CAAGGAAGTACAAAGCCAAATGCCAGGGTCAGAGACCTGCTTGTTGTCAT 1028

Qy   899  GTGATCGGCGGCGACTTTATGCAGAGAGACCATGGGGTCGTCCTAGTGCCAGCCTATATATTTATACTTC 958
          || | | | ||||  |   |  |  ||  |||  | || || |||| || ||| |  ||  | | |
Db   1029 GCCCTTGGAGATCTATTTCAGAGTATCCAGTGTGCTAATTACTGTTAGTGCACTTGATTAATCTTACTTT 1088

Qy   959  AAGCCAGAAATACAAGTATGTAGAGACCCCATCTCCACCTTCGTCTTCCCATCCTCCTGGGTCCTG 1018
          |||| ||  | || ||| | ||| || ||  |||| | || || ||||| | |||||
Db   1089 AAGGCAGAGTATAAAATAGGCGACCCAATCTGCAGAATTCATCATCTTTCCAACCTGGTCTTG 1148

Qy   1019 GGGACAAACCTTGACCCATCCTGAGAGATGTGATCCTGGTTGTTCATGGAAGGACCCCCAAG 1078
          |  | | ||  | ||  ||| |   | | | | || |  | || ||  || || || ||
Db   1149 GCCACCACCATCACTATCTTAAAGGACTTCTCCATCTTACTCATGGAAAGTGTGCCAAAG 1208

Qy   1079 GGCGTTGACTTCACAGCCTGTTCGTGATCTGCTGCTGTGCTGTTGGAGGGGGTAGAAGCCCTG 1138
          |  |||| | |  ||  || |   |  || | || |||| ||  ||| |   |||  | |||
Db   1209 AGCCTGAATTACAGTGTGTGAAAGAGCTTATTTTAGCAGTGCGACCGGGTCGTGCTGTGTG 1268

Qy   1139 CACAGGCCTGCATATCTCGGGCACTGACGGTGGCCCAGCCGTTCGTGCTGTCCACATCGCC 1198
          ||||  |||| |  ||    | |  | ||     ||||| ||||    ||||||  || | |
Db   1269 CACACCCTGCACATCTGGGTCTCTACAACAATGAAATCAAGTAATTCTCTCAGCTCATGTTGCT 1328

Qy   1199 ATTGCTCAGAATACAGAGCGGGACGACCCCAGCTGTGCTGAAGACAGCAGCCGCGCCTCCAAGGG 1258
          | | || ||      || |     | |  ||||  | |  |
Db   1329 ACAGCAGCCATGCGGGACGCCAAGTGGTTCGGGAGAGAAAATTGCTAAAGCCCTTAGCAAA 1388

Qy   1259 AAGTTCCACTTCCACACCGTGACCCATCCAGATCGAGGGACTACTGGCAGACATGAAGGAC 1318
          | |  ||| |  |   |   || ||  | | | | || || | | ||  || || |
Db   1389 AGCTTTACGATGCAACTCACTCACCATTCAGAGAGAATCTCCAGTTGACCAGACCCGAC 1448

Qy   1319 TGTCAGGCATGCCAGGGCCCCTCAGACTGACT 1350
          |||| |||| |   ||||| | ||||| |
Db   1449 TGCCTTTTCTGTGAAGACCCTGTGACTAGCT 1480
```

AC   ABV29801
XX
DT   16-SEP-2002 (first entry)

RESULT 12
ID   ABV29801 standard; cDNA; 2297 BP.
XX
AC   ABV29801;
XX

CC   The invention relates to an isolated nucleic acid molecule (I) comprising
CC   a nucleotide sequence given in Tables 1-9 (ABV00010-ABV62213) of the
CC   specification or its complement. (I) is useful for: (I) is useful for:
CC   (a) assessing whether a patient is afflicted with prostate cancer;
CC   (b) monitoring the progression of prostate cancer in a patient;
CC   (c) assessing the efficacy of a test compound to inhibit prostate
CC   cancer in a patient;
CC   (d) assessing the efficacy of a therapy for inhibiting prostate cancer
CC   in a patient;
CC   (e) selecting a composition for inhibiting prostate cancer in a patient;
CC   (f) assessing the prostate cell carcinogenic potential of a compound;
CC   (g) determining whether prostate cancer has metastasized in a patient;
CC   (h) assessing the aggressiveness or indolence of prostate cancer in a
CC   patient;
CC   (I) is also useful as a pharmacodyanamic or pharmacogenomic marker.
XX
SQ   Sequence 2297 BP; 692 A; 506 C; 491 G; 601 T; 7 other;

Query Match   19.1%; Score 309.6; DB 23; Length 2297;
Best Local Similarity   57.9%; Pred: No. 7.3e-54;
Matches 574; Conservative 0; Mismatches 409; Indels 9; Gaps 1;

```
Qy   359  GAGCTGGGCTGCCCAGAGCAACCATCACTGCCCATGCTCAGAAGGGTCCTGACAGTCACTGT 418
          |||||| |
Db   498  GAGCTGCAGTCAGGAGGCATGTACCACTGCCACACAGTGGCCTCCAAGCCCACACAGAAAAGGGG 557

Qy   419  GACCCCAAGAGAAGGGAAGGCCCAGCGTGATGTAGGCCTCTGCCATCTGCCTGTTG 478
          || |  ||
Db   558  GCGAATGAGTACGCCCTATGCCAAGTGGAAAACTCTGTTGCTTCAGCAAATATGCTTCATT 617

Qy   479  TTCATGATTGCGAGAAGTCGTTGGTTTGGTGGTCACACAACTTGGCTTGCTGCGTCATGACTGAC 538
          |||| ||| || ||
Db   618  TTCATGATTGCAGAGGTCGTGGGTGGGCACACAACTTGGCTTGCTGCGTCATGACTTGTCACAGAT 677

Qy   539  GCAGACAACCTGCTCACTGACTTTGCCAGCAATGCTCATCAGCCTCTTCTCCCTCGGATG 598
          |||| ||
Db   678  GCTGCCAACTTCTTAATTGACCTGACCAGTTTTCCTTCTCAGTCTCTTCTCTCTCGGTTG 737

Qy   599  TCCTCCCCGGCCAGCCACCAAGACCCATGAAACTTTGGCTTGCTGCAGAGAAGCTGAGGATCTTGGGA 658
          |  | |
Db   738  TCATCGAAGCCTCCCTCTTAAGCGGCGTGACATTTGGATGGCACCGAGCCAGAGAATCCTTGGT 797
```

```
RESULT 13
AAF27714
ID   AAF27714 standard; cDNA; 2823 BP.
XX
AC   AAF27714;
XX
DT   28-MAR-2001  (first entry)
XX
DE   Human transport protein TPP1-14 coding sequence.
XX
KW   Human; transport protein; TPP1; transport disorder; metabolic disorder;
KW   neurological disorder; cardiovascular disorder; reproductive disorder;
KW   immune disorder; cancer; ss.
XX
OS   Homo sapiens.
XX
PN   WO200078953-A2.
XX
PD   28-DEC-2000.
XX
PF   16-JUN-2000; 2000MO-US16668.
XX
PR   17-JUN-1999; 99US-0139923.
PR   10-AUG-1999;   99US-0148177.
PR   18-AUG-1999;   99US-0149357.
PR   28-OCT-1999;   99US-0162287.
XX
PA   (INCY-) INCYTE GENOMICS INC.
XX
PI   Lal P, Yang J, Yue H, Hillman JL, Tang YT, Bandman O, Burford N;
PI   Baughn MR, Azimzai Y, Lu DAM, Au-Young J, Patterson C;
XX
DR   WPI; 2001-041424/05.
DR   P-PSDB; AAB60094.
XX
PT   The present invention provides the protein and coding sequences for 43
PT   novel human transport proteins (designated TPP1s). These can be used in
CC   the diagnosis and treatment of transport, metabolic, neurological,
CC   reproductive, cardiovascular and immune disorders, and cell proliferative
CC   disorders such as cancer.
XX
PS   Claim 5; Page 148-149; 165pp; English.
XX
PT   Isolated polypeptide with a human transport protein sequence is useful
PT   for the diagnosis, prevention and treatment of disorders associated
PT   with the immune, reproductive and cardiovascular systems -
XX
SQ   Sequence 2823 BP; 838 A; 599 C; 621 G; 765 T; 0 other;
```

```
Query Match    19.1%;  Score 309.6;  DB 22;  Length 2823;
Best Local Similarity   57.9%;
Matches  574;  Conservative   0;  Mismatches 409;  Indels   9;  Gaps   1;
Pred. No. 7.7e-54;
```

First alignment block:

```
Qy   659 GCCCTGGTCTCTGTACTGTCCATCTCGGTCGTTGCGGTAGTCAGTGGTGTACTGGCTGTG 718
         ||||| |||||   ||   | ||| | |||||| |  | ||||||||||| |||||||
Db   798 GCCCTGCTCTCCATCCTGTGCAATCTGGGTGGTGGTGACTGGCGTCGTTAGTGTACCTGGCATGT 857

Qy   719 GAGCGGCTGATCTCTGGGGACTATGAAATTGACGGGGGGGACCATGCTGATCACGTCGGGC 778
         |||| ||| |  | ||  | || |  ||  |  | | || | | |||| | | ||| || |
Db   858 GAGCGCCTGCTCCATCCTGTGCAATCTGGGTGGTGGTGACTGGCGTCGTTAGTGTACCTGGCATGT 917

Qy   779 TCGCGCTGTTGGCTGTGAACATCATAAATTGGGGTTGACCTTCACCAGTCTGGGCCATGGGCAC 838
Db   918 TCGGCCAGTGGCGGCGGCCAACATTGTTACTAACTGTGTGGTTTGCACCAGAGATGCCCTTGGCCAC 977

Qy   839 AGCCACGGCACCAACCAACCAGGGAGGAGGAAACCCCAGCGTCCGAGCTGCCCTTCATCCAT 898
Db   978 AATCA------CAAGGAAGTACAAAGCCAAATGCCAGCGTCAGAGCGTGCTTTTGTGCAT 1028

Qy   899 GTGATCGGCGACTTTATGCAGAGAGCATGGGGTGTCCTAGTGGTCAGCACCCTATATTTTATACTTC 958
Db  1029 GCCCTTGGAGATCTATTTCAGAGTATCAGTGTGCTAATTAGTGCACTTATTATTCTACTTT 1088

Qy   959 AAGCCAGAATACAAGTATGTAGACCCCATCTCGCTCTTCTCCCATCCTCTGGTCCTG 1018
Db  1089 AAGCCAGAGTATAAAATAGCCGACCCAATCTGCACATTCATTCTTTTCCATCCTGGTCTTG 1148

Qy  1019 GGGACAACCTTGACCATCCTGAGAGATGTGATCCTGGTGTTGATGGAAGGAGGACCCCAAG 1078
Db  1149 GCCAGCACCATCACTACTTAAAGAGACTTCTCCATCTTACCATGTGAAGGTGTGCCAAAG 1208

Qy  1079 GGGGTTGACTTCACAGCTGTTCCTGTCTCCTGTCGGGTGGGAGGGGGTAGAAGCCCTG 1138
Db  1209 AGCCTGAATTACAGTGGGTGTGAAAGAGCTTATTTTACCAGTGCACGGGGTGCTGTCTGTG 1268

Qy  1139 CACAGCCTGCACTATATCTGGGCACTGTGGCCCCAGCCTGTTCTGTCTGTGTCGTCACATCGCC 1198
Db  1269 CACAGCCTGCACATCTGGTCTCTTAACATGAATCAAGTAATTCTCTCTCAGCTCATGTTGCT 1328

Qy  1199 ATTGCTCAGAATACAGACGCCCAGGGCTGCTGAAGAAGCAGCAGCCGCGCCTCCAAGGG 1258
Db  1329 ACAGCACCCAGCGCGGGACAGCCAAGTGGTTCGGAGAGAAAATTGTCTAAAGCCCTTAGCAAA 1388

Qy  1259 AAGTTCCACTTCACACCGTGACCATCCAGATCGAGGACTACTCGGAGGAGACAATGAAGGAC 1318
Db  1389 AGCTTTACGATGCACACTCACTCACCATTCAGATGGAAATCTCCAGTTGACCAGGAGACCCGAC 1448

Qy  1319 TGTCAGGCATGCCAGGGCCCCCTCAGACTGACT 1350
Db  1449 TGCCTTTTCTGTGTGAAGACACCCTGTGACTAGCT 1480
```

Second alignment block:

```
Qy   359 GAGCTGGCTGCCCAGAGCAACCATCACTCACCATGCCATCTCAGAAGGGTCACGTCACTGT 418
Db   439 GAGCTGGACTCAGCAGCAGTACCCATCCACCAGTGATGATCAACGTTTCCAGC 498

Qy   419 GACCCCAAGAGAAGGGGAAGGCCCAGCGGCCAGTCGTATGTAGCTCTGCCATCTGCCTGTG 478
Db   499 GCGAATGAGTAGCCCTATCCCAAGTGGGAAAACTCGTTCTGCTTCAGCAATATGCTTCATT 558

Qy   479 TTCATGATCGGAGAAGTCGTTGGTGGGTACCTGGCACACAGCTTGGCTGCTGTCATGACTGAC 538
Db   559 TTCATGATTGCAGAGGTCGTGGGTGGGCACAGGTGGAGTCGTTGGGCACAGAT 618

Qy   539 GCAGCACACCTGCTCACTGCACTTTGCCAGCCATGGCCATTGCTGGAAGACCTCTCTCCCTCTGGA 598
Db   619 GCTGCCCACCTCTTAATTGACCTGACCAGTTTCCTGCTACCAGTCTCTTCTCCCTGGTGTG 678

Qy   599 TCCTCCCGGCCAGCCACCAAGAACCATGAACTTTGGCGTGACAGGGGTGTGGTGACCGTGTG 658
Db   679 TCATCGAAGGCCTCCCTCTCAAGCGGCGCCTGACCATTTGGAGTGGCACCGGAGAGATCTTGGGA 738
```

Qy 1019 GGGACAAACCTTGACCAATCCTGAGAGAGGATGGATCCTGCTGTGTTGTTGCATGGAAGGAGACCCCCAAG 1078
        || ||| |||||||||||| |||
Db 1090 GCCAGCACCAATCACTCATATCTTAAAGGACTTCCCATCTTACTCATGGAAGGTGTGCCAAAG 1149

Qy 1079 GGCGGTTGACTTCACAGCGTGTTCGTTCGTGATCGTCGGTGGGAGGGGGTAGAAGCCCTG 1138
        ||||| ||||||||||| |||||||| |||
Db 1090 GCCAGCACCAATCACTCATATCTTAAAGGACTTCCCATCTTACTCATGGAAGGTGTGCCAAAG 1149

Qy 1139 CACAGCCTGCATATCTGGGCACTGAGGGTGCCCAGCCGTTCGTTCGTGTCGTCCACATCGCC 1198
        |||||||| |||||| ||| |||||| ||| | ||||||||||||||| ||||
Db 1150 AGCCTGAATTACAGTGGTGTGAAAGAGCTTATTTTACAGTGCACGGGTGCTCATGTTGCT 1209

Qy 1199 ATTGCTCAGAATACAGACGCGCCAGGCGTGTGCTGAAGCAGCACGCGCCGCTCCAAGG 1258
        ||||||| ||||||||| ||||||| |||
Db 1270 ACAGCAGCCACGCGGGACAGCCAAGTGGTTCGGACAGGAAATTGCTAAAGCCCTTAGCAAA 1329

Qy 1259 AAGTTCCACTTCCACACCGTGACCATCCACAGATCGGAGGACTACTCGGAGGACAATGAAGGAC 1318
        | ||||| || || |||| |||| |||
Db 1330 AGCTTTACGATGCACTCACTGCACCATTCAGATGGGAATCTCCAGTTGACCAGGACCCCGAC 1389

Qy 1319 TGTCAGGCATCCAGGCGCCCCCTCAGACTGACT 1350
        ||||||| ||| |||
Db 1390 TGCCTTTTCTGTGAAGACCCTGTGACTAGCT 1421

RESULT 14

AAD28259

ID   AAD28259 standard; cDNA; 2852 BP.
XX
AC   AAD28259;
XX
DT   22-APR-2002 (first entry)
XX
DE   Human pancreatic tumour protein full-length cDNA, 981418.1.
XX
KW   Human; pancreatic tumour protein; pancreatic cancer; gene therapy;
KW   vaccine; ss.
XX
OS   Homo sapiens.
XX
FH   Key   Location/Qualifiers
FT   CDS   457..1419
FT   /*tag= a
FT   /product= "Human pancreatic tumour full length
FT   protein, 981418.1"
XX
PN   WO200194409-A2.
XX
PD   13-DEC-2001.
XX
PF   31-MAY-2001; 2001WO-US18003.
XX
PR   07-JUN-2000; 2000US-210329P.
PR   14-NOV-2000; 2000US-248980P.
PR   15-MAY-2001; 2001US-291197P.
XX
PA   (CORI-) CORIXA CORP.
XX
PI   Hirst SK, Harlocker SL, Dillon DC, Kalos MD;
XX
DR   WPI; 2002-154565/20.
DR   P-PSDB; AAB17562.
XX
PS   Claim 1; Page 109; 115pp; English.
XX
PT   Pancreatic tumour polypeptide and polynucleotide useful in diagnosis,
PT   prevention and/or treatment of cancer, especially pancreatic cancer -
XX
CC   The invention relates to human pancreatic tumour polypeptides and
CC   nucleic acid molecules encoding such polypeptides. The invention
CC   also relates to compositions and methods for the diagnosis,
CC   prevention and therapy of cancer, particularly pancreatic cancer.
CC
CC   Sequences of the invention are used in gene therapy and in vaccines.
CC   The present sequence is human pancreatic tumour protein full-length
CC   cDNA.
XX
SQ   Sequence 2852 BP; 863 A; 601 C; 620 G; 768 T; 0 other;

Query Match   19.1%;   Score 309.6; DB 24;   Length 2852;
Best Local Similarity   57.9%;   Pred. No. 7.7e-54;
Matches   574;   Conservative   0;   Mismatches 409;   Indels   9;   Gaps   1;

Qy 359 GAGCTGGCTGCCCAGAGCAACACTCACTGCCCATGCTCAGAAAGGGTCCTGACAGTCACTGT 418
         ||||||||| ||||| |||
Db 679 TCATCGAAGCCTCCCTCTTAAGCCGGCTGACATTTGGATGGCACCGGCAGAGAGGACCCTTGCT 738

Qy 419 GACCCCAAGAAGGGAGGGGAAGGCCCAGCCGGCGTGTATGTTAGCCTTGCCATCTTGCCTGTTG 478
         |||||| ||||||| |||
Db 739 GCCCTGCTCTCCATCCCTGTGCAATCTGGGCTGTGGTGTCGCGGGGGGGTACTGGTGGCTGTG 798

Qy 479 TTCATGATTGCAGAGGTCGTTGGTGGGTACCTGGCACACAGCGTTCGCTTCAGCAATATCCTTCATT 538
         ||||||||| |||||| |||
Db 799 GAGCGGCCCTGGCTGTTCATCCGATTACCCAGATCCAGCGACTGTGATGAATCATCGTTTGCAGC 858

Qy 539 GCAGCACACCTGCTCACTGACATTTGTCAGCATGCTCATCAGCCTCTTCTCCCTCTGGATG 598
         |||||| ||||| |||
Db 859 TGCCGCCAGTGGCGCGCCCAACCATTGTTACTAACCTGTGCGGTTTGACCGAGAGCTCTGGCCAC 918

Qy 599 TCCTCCCGGCCACCCAGGCCCACCAAGACCATGAAGCTTTGGGCTGGCAGAGAGCTTGGGA 658
         ||||| ||||||| |||
Db 919 AATCA------------CAAGGAAGTAACAAGCCAATGCCAGCGTCAGAGAGCTGACTTTGTGCAT 969

Qy 659 GCCCTGGTCTCTGTTGTCTACTGTCCATCTTGGGTCTCGGGGGGGGGGGTACTGGTGGCTGTG 718
         ||||| ||||||| |||
Db 970 GCCCTTGGAGATGTCATTCCAGAGTACTTCAGTGTGCTAATTAGTGCACTTATTATCTACTTT 1029

Qy 719 GACGGGCTGATCTCTGGGGACTATGAAATTGACGGGGGGACCATGCTGATCAGTGCGGGC 778
         ||||| ||||||| |||
Db 1030 AAGCCAGAATACAAAGTATGTAGAGCCGATGACGGTGGCACTGAGGACTTATTTTACAGTGCACGGGTGCTC 1089

Qy 779 TCGCTGTGCTTGGGCTGGACATCATTAAATGGGTTGACCCTTCACCAGTTCCACTTTATACTTC 838
         ||||| ||||||| |||
Db 1090 GCCCTTGGAGATGTCATTCCAGAGTACTTCAGTGTGCTAATTAGTGCACTTATTATCTACTTT 1149

Qy 839 AGGCCACGGCACCACCAGGAGAGGAGAACCCAGCGTCCGAGCTGCCTTCATCGCAT 898
         ||||| ||||||| |||
Db 1150 AGCCTGAATTACAGTGGTGTGAAAGAGCTTATTTTACAGTGCACGGGTGCTCATGTTGCT 1209

Qy 899 GTGATCGGCGACTTTATCGAGACCATGGGTGTGTCCTCAGTGGCAGCCTATATTTTTATACTTC 958
         ||||| ||||||| |||
Db 970 GCCCTTGGAGATGTCATTCCAGAGTACTTCAGTGTGCTAATTAGTGCACTTATTATCTACTTT 1029

Qy 959 AAGCCAGAATACAAAGTATGTAGAGCCGATGACGGTGGCACTGAGGACTTATTTTACAGTGCACGGGTGCTC 1018
         ||||| ||||||| |||
Db 1018 AAGCCAGAATACAAAGTATGTAGAGCCGATGACGGTGGCACTGAGGACTTATTTTACAGTGCACGGGTGCTC 1089

Qy 1079 GGCGGTTGACTTCACAGCGTGTTCGTTCGTGATCGTCGGTGGGAGGGGGTAGAAGCCCTG 1138
         ||||| ||||||| |||
Db 1150 AGCCTGAATTACAGTGGTGTGAAAGAGCTTATTTTACAGTGCACGGGTGCTCATGTTGCT 1209

Qy 1019 GGGACAAACCTTGACCAATCCTGAGAGAGGATGGATCCTGCTGTGTTGTTGCATGGAAGGAGACCCCCAAG 1078
         ||||| ||||||| |||
Db 1210 CACAGCCTGCATATCTGGGCACTGAGGGTGCCCAGCCGTTCGTTCGTGTCGTCCACATCGCC 1198

Qy 1139 CACAGCCTGCATATCTGGGCACTGAGGGTGCCCAGCCGTTCGTTCGTGTCGTCCACATCGCC 1198
         ||||||| |||||| |||
Db 1270 ACAGCAGCCACGCGGGACAGCCAAGTGGTTCGGACAGGAAATTGCTAAAGCCCTTAGCAAA 1329

Qy 1199 ATTGCTCAGAATACAGACGCGCCAGGCGTGTGCTGAAGCAGCACGCGCCGCTCCAAGG 1258
         ||||| ||||||| |||
Db 1259 AAGTTCCACTTCCACACCGTGACCATCCACAGATCGGAGGACTACTCGGAGGACAATGAAGGAC 1318

Qy 1259 AAGTTCCACTTCCACACCGTGACCATCCACAGATCGGAGGACTACTCGGAGGACAATGAAGGAC 1318
         ||| |||||| |||
Db 1330 AGCTTTACGATGCACTCACTGCACCATTCAGATGGGAATCTCCAGTTGACCAGGACCCCGAC 1389

```
Qy      1319 TGTCAGGGCATGCCAGGGCCCCTCAGACTGACT 1350
              |||   || |   | ||||   |  ||||||  ||
Db      1390 TGCCTTTCTGTGAAGACCCCTGTGACTAGCT 1421

RESULT 15
ID      ABK71913 standard; cDNA; 1623 BP.
XX
AC      ABK71913;
XX
DT      30-JUL-2002  (first entry)
XX
DE      Human cDNA encoding zinc transporter-like protein NOV2b.
XX
KW      Human; ss; gene; NOVX; developmental disorder; endocrine disorder;
KW      vascular disorder; infectious disease; endocrine disorders; stroke;
KW      neurodegenerative disorder; Alzheimer's disease; acute brain injury;
KW      central nervous system disorder; depression; lung disorder;
KW      reproductive disorder; tissue disorder; thrombocytopenia; migraine;
KW      angiogenesis; asthma; x-linked severe combined immunodeficiency;
KW      inflammation; autoimmune disorder; immune disorder; blood disorder;
KW      haematopoietic disorder; gastrointestinal disease; respiratory disorder;
KW      hepatitis; fertility; hypertension; arteriosclerosis; ischaemia;
KW      rheumatoid arthritis; Grave's disease; wound healing.
XX
OS      Homo sapiens.
XX
PN      WO20024733-A2.
XX
PD      28-MAR-2002.
XX
PF      17-SEP-2001; 2001WO-US29115.
XX
PR      15-SEP-2000; 2000US-232675P.
PR      15-SEP-2000; 2000US-232676P.
PR      15-SEP-2000; 2000US-232679P.
PR      18-SEP-2000; 2000US-233382P.
PR      18-SEP-2000; 2000US-233402P.
PR      19-SEP-2000; 2000US-233521P.
PR      19-SEP-2000; 2000US-233522P.
PR      19-SEP-2000; 2000US-233801P.
PR      20-SEP-2000; 2000US-233960P.
PR      06-OCT-2000; 2000US-238398P.
PR      13-OCT-2000; 2000US-240284P.
PR      13-OCT-2000; 2000US-240498P.
PR      11-JAN-2001; 2000US-260973P.
PR      26-JAN-2001; 2001US-264274P.
PR      09-MAR-2001; 2001US-274862P.
XX
PA      (CURA-) CURAGEN CORP.
XX
PI      Mishra VS,   Syptek KA,   Taupier RJ,   Vernet CAM,   Colman SD,   Gorman L;
PI      Tchernev VT, Malyankar UM, Shenoy S,  Tcherrev VT,  Padigaru M;
PI      Patturajan M, Burgess CE, Smithson G, Millet I, Peyman JA;
PI      Stone D, Gunther E, Ellerman K;
XX
PI      WPI; 2002-383182/41.
PI      P-PSDB; ABG60225.
XX
PT      New cytoplasmic, nuclear, membrane bound and secreted NOVX
PT      polypeptides, useful for treating cancers and tumours, lung disorders,
PT      haematopoietic disorders, autoimmune diseases and immune disorders -
XX
PS      Claim 9; Page 20-21; 210pp; English.
XX
CC      The invention relates to an isolated NOVX polypeptide selected from
CC      NOV1a, NOV1b, NOV1ac, NOV2a, NOV2b, NOV2c, NOV3a, NOV3b, NOV4a, NOV4b,
CC      NOV5a, NOV5b or NOV6-NOV9 polypeptides; their mature form or variant.
CC      Also included are a nucleic acid encoding a NOVX protein or variant;
CC      a vector comprising the nucleic acid; a cell comprising the vector;
CC      an anti-NOVX antibody; and identifying agents that modulate the
CC      expression or activity of NOVX. NOVX, the nucleic acid, antibody and
CC      modulators are useful in the diagnosis, treatment or prevention of
CC      developmental disorders, endocrine disorders, vascular disorders,
CC      infectious disease, endocrine disorders, cancer, neurodegenerative
CC      disorders, Parkinson's disease, Huntington's disease, multiple
CC      Alzheimer's disease, Parkinson's disease, neurodegenerative disorders
CC      sclerosis and amyotrophic lateral sclerosis), acute brain injury (e.g.
CC      stroke, head injury and cerebral palsy), central nervous system
CC      (e.g. depression, epilepsy and schizophrenia), lung disorders,
CC      reproductive disorders, disorders affecting carbohydrate metabolism (e.g.
CC      galactosaemia and hereditary fructose intolerance), tissue disorders
CC      (e.g. Wiskott-Aldrich syndrome, thrombocytopenia, night blindness and
CC      Pick's disease), disorders linked to abnormal angiogenesis, asthma,
CC      azoospermia, learning disabilities, facial dysmorphism, autoimmune
CC      encephalomyelitis, X-linked severe combined immunodeficiency, seizures,
CC      migraines, inflammation, autoimmune disorders, disorders affecting sleep,
CC      appetite, thermoregulation, pain, perception, hormone secretion and
CC      sexual behaviour, immune disorders, haematopoietic disorders or other
CC      disorders related to cell signal processing and metabolic pathway
CC      modulation, gastrointestinal diseases, respiratory disorders, blood
CC      disorders, hepatitis, trauma, regeneration, viral, bacterial or parasitic
CC      infections, hyper- or hypo-thyroidism, endometriosis, fertility,
CC      hypertension, arteriosclerosis, ischaemia, haemolytic anaemia, Werner
CC      syndrome, rheumatoid arthritis, Grave's disease, wound healing, X-linked
CC      mental retardation, psychotic and neurological disorders and neuronal
CC      degeneration. The present sequence encodes a NOVX protein.
XX
SQ      Sequence 1623 BP; 437 A; 380 C; 358 G; 448 T; 0 other;

Query Match             16.6%;  Score 267.8;  DB 24;  Length 1623;
Best Local Similarity   55.2%;  Pred. No. 2.4e-45;
Matches  549;  Conservative    0;  Mismatches 437;  Indels   9;  Gaps  1;


Qy       359 GAGCTGGCTGCCCAGAGCAACCATCACTCGCCATGCTTCAGAAGGGTGCTTGACAGTCACTGT 418
             |||   | ||| || |  |  | ||||||||  ||| |
Db       421 GAGCTGGAGTTCAGGAGGGCATGTACCACTGCCACACTGCCACCACCACACAGAAAAGGGG 480

Qy       419 GACCCCAAGAAAGGGGAGAAGGGCCCAGCCCACCTGTGTATATTACCCTTCGCCAACTCTGTTG 478
             | | |||||||| |   |   | ||| |  |
Db       481 GCGAATGAGTACGGCCTTATGCCAAGTGGAAAACTCGTTCTGCTTCACCAAATATGCTTCATT 540

Qy       479 TTCATGATCGGAGAAGTCGTTGGTGGGGTACCTGGCCACACAGCTTGGCTGCTGACTGAC 538
             |||||| ||| |  |   ||||| || ||  |||  |
Db       541 TTCATGATTCCAGAAGGTCGTGGGGTGGGCACATTGCTGGGGAGTCTTGCTGTTGTCACAGAT 600

Qy       539 GCAGCACACCTGCTCACTGACTTTGCCAGCATGGTACGGGGGTACTGGTGGTGTACCTGGCTG 598
             | || |  |  || | ||||   | |||| |   |  |  |   ||| |
Db       601 GCTGCCCACCTCTTAATTGACCTGACCAGTTTCCTGCTCAGTCTCTCTCCCTGTGGTTG 660

Qy       599 TCCTCCCGGCCCAGCCACCCAAGAACACTTTGGCTGGAGAGAGCTGAGAAGATCTTTGGGA 658
             |  || | ||  | |  | |    |   | |||||||| |   |  | ||||||
Db       661 TCATCGAAGCCCTCCCCTCTAAGCGGCTGACATTTGGATGGCACCGACCACAAGGTTTTATTT 720

Qy       659 GCCCCTCGGTCTCTGTACTGTCCCATCTGGGTCGTGACGGGGGTACTGGTGTGTACCTGGCGTG 718
             | || | ||| | ||| ||| | || ||| |||| |   |  ||  || |  ||   |||
Db       721 AGCATTTTAATCTCTCATCACCCTGGTTGTGTGGCTGTGCTAGTGTACCTGGCATGT 780

Qy       719 GAGCGGCTGCGATCTCTGGGGACTATGAAAATTGACGGGGGACCATGCTGATCACGTCGGGGC 778
             | || |    |   || |  |  |    |  |  |  | |||  |    |   |  |  |
Db       781 GAGCGCGCCTGCTGCTGTATCCTGATTACCAGATCCAGGCGACTGTGAGTCATCGTTTCCAGC 840

Qy       779 TGCGCTGTGTGGCCGTGTGAAACATCATAATGGGGTTGACCCTTCACCAGTCTGGCCATGGGCAC 838
             ||||| |||||  || |     | | ||  || |    | |   |   |||||| |  ||| |
Db       841 TGCGCAGTGCGGCGCCCGCTAAGACAATTTGTTCTCTCTTTCAGACTAACTGTGGTTTTGCAC 900

Qy       839 AGCCACAGGCACCACCAACCAGCAGGAGGA------------GAACCCCAGGCGTCCGACGTGCC 889
             |  |||| |     | | | ||| || || |           ||||||||| ||| | ||||
Db       901 CAGAGATGCCTTGGCCGCAATCACAAGGAGTACAAGCCAATGCCAGCGTGCGACAGCTGCT 960

Qy       890 TTCATCCATGTGATCGGCCGACTTTATGCAGAGCACATGGGTGTTGGTGCCTCAGTGGCAGCCTATATT 949
             | |  | |    |   |      ||||| ||   |  |   | |   |  |||| |  | | ||
Db       961 TTTGTTGCATGCCTTGGAGATCTATTTCAGATTATGGCTATGTGTTGGCTAATTAGTTAGCACTTATT 1020

Qy       950 TTATACTTCAAGCCAGAATACAAGTATGTAGAGAGACCCCATCGTGCACCTTCGTCGTCTTCCATC 1009
             |  || |   ||| |  |  |  |  ||   |  ||||  |  || ||||| | |||| |  |||
```

```
Db  1021 ATCTACTTTTAAGCCAGAGTATAAAATAGCCGACCCAATCTGCACATTCATCTTTTCCATC 1080
              |||||  |||||||||| || |  |  ||||||||||||    || ||| ||| |||||
Qy  1010 CTGGTCCTGGGGACAACCTTGACCATCCTGAGAGATGTGATCCTGGGTGTTGATGGAAGGG 1069
         |||||| ||| | ||||| |  ||  ||  |   || |  | || |   || |||||||||
Db  1081 CTGGTCTTGGGCAGCACCATCTCTATCTTAAAGGACTTCTTCTTCTTACTCATTGGAAGGT 1140

Qy  1070 ACCCCCAAGGGCGTTGACTTCACAGCTGTTCGTGATCTGCTGCTGTCGGTGGAGGGGGTA 1129
              ||| || | ||||| |  |  |  ||  || ||   |  ||  | || |||||||
Db  1141 GTGCCAAAGAGGCCTGAATTACAGTGGGTGTGAAAGAGCTTATTTTATCAGTCGACGGGGTG 1200

Qy  1130 GAAGCCCTGCACAGCCTGCATATTCTGGGCACTGACGGTGGCCCAGCCTGTTCTGTCTGTC 1189
             |||||||||||||||  |  |    || || ||  ||| |  |||||| ||| ||||
Db  1201 CTGTCTCTGTGCACAGCCTGCACANTCTGGTCTCTAACAANTGAAGTAATTCTCTCAGCT 1260

Qy  1190 CACATCGCCATTGCTCAGAAATACAGACGCCCAGGCTGTGCTGANAGACAGCCAGCCGC 1249
              |||  |  || ||  || |  ||     ||| |  ||  || |  ||||||
Db  1261 CATGTTGCTACAGCAGCCGCGGGACAGCCAAGTGGTTCGGAGAGAAATTGCTAAAGCC 1320

Qy  1250 CTCCAAGGGAAGTTCCACTTCCACACCGTGACCATCCAGATCGGAGGACTACTCGGAGGAC 1309
              |  |  |  ||||||    | ||||||||||  |
Db  1321 CTTAGCAAAAAGCTTTACGATGCACTCACCATTCAGATGGAATCTCCAGTTGACCAG 1380

Qy  1310 ATGAAGGACTGTCAGGCAATGCCAGGGCGCCCTCAGA 1344
              |||||| ||||| | |||| |||| |||||||||
Db  1381 GACCCGACTGCCTTTTCTGTGAAGACCCCTGTGA 1415
```

OM nucleic - nucleic search, using sw model

GenCore version 5.1.4_p5_4578
Copyright (c) 1993 - 2003 Compugen Ltd.

Run on:　　　March 29, 2003, 16:20:14 ; Search time 2965.84 Seconds
　　　　　　　　(without alignments)
　　　　　　　　8829.908 Million cell updates/sec

Title:　　　　　US-09-691-219-1
Perfect score:　1617
Sequence:　　　1 cgcggtgaacagcacgaggg.........aaaaaaaaaaaaaaatgtt 1617

Scoring table:　IDENTITY_NUC
　　　　　　　　Gapop 10.0 , Gapext 1.0

Searched:　　　16154066 seqs, 8097743376 residues

Total number of hits satisfying chosen parameters:　　32308132

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
　　　　　　　　　Maximum Match 100%
　　　　　　　　　Listing first 45 summaries

Database :
```
        EST:*
    1:  em_estba:*
    2:  em_esthum:*
    3:  em_estin:*
    4:  em_estmu:*
    5:  em_estov:*
    6:  em_estpl:*
    7:  em_estro:*
    8:  em_htc:*
    9:  em_est1:*
   10:  gb_est2:*
   11:  gb_htc:*
   12:  gb_est3:*
   13:  gb_est4:*
   14:  gb_est5:*
   15:  em_estun:*
   16:  em_estom:*
   17:  gb_gss:*
   18:  em_gss:*
   19:  em_gss_hum:*
   20:  em_gss_inv:*
   21:  em_gss_pln:*
   22:  em_gss_vrt:*
   23:  em_gss_fun:*
   24:  em_gss_mam:*
   25:  em_gss_mus:*
   26:  em_gss_other:*
   27:  em_gss_pro:*
       em_gss_rod:*
```

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

```
Result                   %    Query
No.     Score   Match   Length   DB   ID            Description
-----   -----   -----   ------   --   --            -----------
  1     590     36.5     732      9   AU135613      AU135613 AU135613
  2     565.6   35.0    1627     14   BM918933      BM918933 AGENCOURT
  3     521     32.2     783      9   AU135737      AU135737 AU135737
  4     510.8   31.6     931     14   BQ219887      BQ219887 AGENCOURT
  5     410.8   25.4     828     12   BE746716      BE746716 601579148
  6     389.2   24.1     394     12   BF509598      BF509598 UI-H-BI4-
```

ALIGNMENTS

RESULT 1
AU135613
AU135613

```
LOCUS       AU135613              732 bp    mRNA    linear
DEFINITION  AU135613 PLACE1 Homo sapiens cDNA clone PLACE1002500 5', mRNA
            sequence.
ACCESSION   AU135613
VERSION     AU135613.1  GI:10996152
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 732)
  AUTHORS   Ota,T., Nishikawa,T., Suzuki,Y., Ishii,S., Saito,K., Kawai,Y.,
            Yamamoto,J., Wakamatsu,A., Nakamura,Y., Nagai,T., Sugano,S. and
            Isogai,T.
  TITLE     HRI human cDNA project
  JOURNAL   Unpublished (2000)
COMMENT     Contact: Takao Isogai
            Genomics Laboratory
            Helix Research Institute
            1532-3 Yana, Kisarazu, Chiba 292-0812, Japan
            Tel: 81-438-52-3975
            Fax: 81-438-52-3986
            Email: genomics@hri.co.jp
            HRI human cDNA project: 5'- & 3'-end one pass sequencing: Helix
            Research Institute; cDNA library construction: Department of
            Virology, Institute of Medical Science, University of Tokyo, and
            Helix Research Institute.
```

7	373.4	23.1	662	12	BE741089	BE741089	601594088
8	325.4	20.1	598	9	AU138346	AU138346	AU138346
9	298.2	18.4	319	9	AI925963	AI925963	wh12g11.x
10	293.6	18.2	599	10	BB620842	BB620842	BB620842
11	271.8	17.2	567	10	BB649221	BB649221	BB649221
12	271.8	16.8	674	9	AL651875	AL651875	AL651875
13	267.4	16.5	524	10	BE666434	BE666434	149858 MA
14	257.8	15.9	608	10	BQ384277	BQ384277	NISC_mn05
15	244.8	15.1	619	13	BI349390	BI349390	dac60g08.
16	242	15.0	373	9	AA684057	AA684057	vm52f11.r
17	240	14.8	664	9	AL631343	AL631343	AL631343
18	238.4	14.7	557	10	BE232116	BE232116	137049 MA
19	235.2	14.5	514	10	BE232050	BE232050	136568 MA
20	223.6	14.1	520	9	AL792631	AL792631	AL792631
21	223.6	13.8	586	13	BJ042242	BJ042242	BJ042242
22	222.2	13.7	646	9	AL782009	AL782009	AL782009
23	218.8	13.5	650	9	AL641096	AL641096	AL641096
24	217	13.4	1004	14	BM925306	BM925306	AGENCOURT
25	211.8	13.1	547	10	BE014623	BE014623	126423 MA
26	211.4	13.1	628	9	AL782327	AL782327	AL782327
27	210.4	13.1	546	13	BJ042222	BJ042222	BJ042222
28	210.4	13.0	627	9	AL799760	AL799760	AL799760
29	209	12.9	478	9	AL792775	AL792775	AL792775
30	208.2	12.9	631	9	AL652511	AL652511	AL652511
31	208.2	12.9	1013	13	BM554028	BM554028	AGENCOURT
32	207.4	12.8	544	14	BQ417284	BQ417284	1k36b01.y
33	204	12.6	561	12	BG087791	BG087791	H3145A06-
34	199.4	12.3	673	13	BI289597	BI289597	UI-R-DK0-
35	196	12.1	452	10	AW147502	AW147502	da03g09.Y
36	192.4	11.9	652	13	BJ038722	BJ038722	BJ038722
37	190.2	11.8	749	13	BJ056206	BJ056206	BJ056206
38	189.4	11.7	524	13	BJ524737	BJ524737	BJ524737
39	189	11.7	638	13	BJ013993	BJ013993	BJ013993
40	188.8	11.7	752	13	BJ028152	BJ028152	BJ028152
41	185.4	11.5	637	13	BM386576	BM386576	UI-R-CN1-
42	181.2	11.2	593	13	BJ062415	BJ062415	BJ062415
43	179	11.1	372	12	BF524462	BF524462	UI-R-C3-t
44	173.6	10.7	573	12	BG609396	BG609396	323225 MA
45	167.8	10.4	851	9	AL558140	AL558140	AL558140

```
FEATURES             Location/Qualifiers
     source          1..732
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="PLACE1002500"
                     /clone_lib="PLACE1"
                     /note="Vector: pME18SFL3"
                     /tissue_type="placenta"
BASE COUNT    155 a    226 c    200 g    146 t    5 others
ORIGIN

Query Match          36.5%;  Score 590;  DB 9;  Length 732;
Best Local Similarity   99.8%;  Pred. No. 5.6e-81;
Matches    601;  Conservative    0;  Mismatches   0;  Indels   1;  Gaps   1;

Qy      966 AATACAAGTANGTAGACCCCATCTGACCTTCGTCTTCCATCCTGGTCCTGGGGGACAA 1025
            |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        1 AATACAAGTANGTAGACCCCATCTGACCTTCGTCTTCCATCCTGGTCCTGGGGGACAA   60

Qy     1026 CCTTGACCATCCTGAGAGATGTGATCCTGGTGTTGTTGATGGAAGGAGACCC...      1085
            |||||||||||||||||||||||||||||||||||||||||||||||||||
Db       61 CCTTGACCATCCTGAGAGATGTGATCCTGGTGTTGTTGATGGAAGGAGACCC...      120

Qy     1086 ACTTCACAGCGTGTTCGTGATCTGTCTGTGTGGGTGAGGGGGTAGAAGCCCTTGCACAGCC 1145
            |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      121 ACTTCACAGCGTGTTCGTGATCTGTCTGTGTGGGTGAGGGGGTAGAAGCCCTTGCACAGCC  180

Qy     1146 TGCATATATCTGGGCACTGACGGTGGCCCCAGCCTGTCTTCTCGTCTCCACATCGCCATTGCTC 1205
            |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      181 TGCATATATCTGGGCACTGACGGTGGCCCCAGCCTGTCTTCTCGTCTCCACATCGCCATTGCTC  240

Qy     1206 AGAATACAGACCCCCAGGCTGTGCTGAAGACAGCCACCAGCCGCCTCCAAGGGAAGTTCC 1265
            |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      241 AGAATACAGACCCCCAGGCTGTGCTGAAGACAGCCACCAGCCGCCTCCAAGGGAAGTTCC  300

Qy     1266 ACTTCACACCGTGACCATCCAGATGAGGGGACTACTCGGAGGACAGAAGGACTGTCAGG 1325
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      301 ACTTCACACCGTGACCATCCAGATGAGGGGACTACTCGGAGGACAGAAGGACTGTCAGG  360

Qy     1326 CATGCCAGGGCCCCTCAGACTACTCGGTCAGCAGGGCATGAAACAGGAC 1385
            |||||||||||||||||||||||||||||||||||||||||||||||||
Db      361 CATGCCAGGGCCCCTCGAGACTACTGACTCTCACCCGGGGCATGAAACAGGAC  420

Qy     1386 CTGCAGGTGGCTGGACTGAGTGTCCCCCAGGGGCCCAGGGCCCTACCCCAGCTG 1445
            ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      421 CTGCAGGTGGCTGGGACTGAGTGTCCCCAGCCGCCCAGGGACTTTGCCTACCCCAGCTG  480

Qy     1446 TGTTATATAAAACCAGGTCCCCTCCCCCGCCCCACTCCAGGAANTGGAGCTCTTCCC- 1504
            |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      481 TGTTATATAAAACAGGTCCCCCTCCTCGACCTCTGCCCACTCCAGGAANTGGAGCTCTTCCCA  540

Qy     1505 AGCCTCCCATCTGACTACAGCCAGGGTGTATAAAGCTAGTGTGACCC 1564
            |||||||||||||||||||||||||||||||||||||||||||||
Db      541 AGCCTCCCATCTGACTACAGCCAGGGTGGGGACTCAGCGGGTATAAAGCTAGTGTGACCC  600

Qy     1565 TG 1566
            ||
Db      601 TG  602
```

RESULT 2

```
BM918933
LOCUS       BM918933                1627 bp    mRNA    linear   EST 12-MAR-2002
DEFINITION  AGENCOURT_6653483 NIH_MGC_120 Homo sapiens cDNA clone IMAGE:5748088
            5', mRNA sequence.
ACCESSION   BM918933
VERSION     BM918933.1  GI:19369312
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 1627)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: Life Technologies, Inc.
            cDNA Library Preparation: Life Technologies, Inc.
            cDNA Library Arrayed by: The I.M.A.G.E. Consortium
            DNA Sequencing by: Agencourt Bioscience Corporation
            Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLAM2775 row: e column: 17
            High quality sequence stop: 543.

FEATURES             Location/Qualifiers
     source          1..1627
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:5748088"
                     /clone_lib="NIH_MGC_120"
                     /note="Organ: pooled pancreas and spleen; Vector:
                     pCMV-SPORT6; Site.1: NotI; Site.2: EcoRV (destroyed); RNA
                     source anonymous pool of spleen and pancreas from 28 yo
                     male. Library is oligo-dT primed and directionally cloned
                     (EcoRV site is destroyed upon cloning). Average insert
                     size 1.5 kb, insert size range 1-2.5 kb. Library is
                     normalized and enriched for full-length clones and was
                     constructed by C. Gruber (Invitrogen). Research Genetics
                     tracking code 025. Note: this is a NIH_MGC Library."
BASE COUNT    294 a    532 c    479 g    317 t    5 others
ORIGIN

Query Match          35.0%;  Score 565.6;  DB 14;  Length 1627;
Best Local Similarity   97.9%;  Pred. No. 2.3e-77;
Matches    605;  Conservative    0;  Mismatches   9;  Indels   4;  Gaps   3;

Qy       20 GGGAGCCGCTTGGCACGCGCGGAGCCCAGGAGCCGAGCTGCAGCCGCAGCCGGAGCCGGAG   79
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        1 GGGAGCCGCTTGGCACGCGCGGAGCCCAGGAGCCGAGCTGCAGCCGCAGCCGGAGCCGGAG   60

Qy       80 CTCAGGGGCCGCAGGAGCCGCGGAGAGTGAGCGCACCTCGCGGGGGCCCTCGGGGTCAGGT  139
            ||||||||| |||||||||||||| |||||||||||||||||||||||||||||||||| |
Db       61 CTCAGGGGCGGCGCAGGAGCCGCGGAGTGAGCGCACCTCGCGGGGGCCCTCGGGGTCAGT  120

Qy      140 GGGTGAGGCGCCACCCGGAGGAGTCCCGGCGCCCCAACTTTCAGGGGCGGCCACTCGGGGCCTG  199
            |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      121 GGGTGAGGCGCCACCCGGAGGAGTCCCGGCGCCCCAACTTTCAGGGGCGGCCACTCGGGGCCTG  180

Qy      200 CGCGGGCTGCCGGGACTCGGCGCGGGGACTGCATGGAGGGCCAAGGAGAGGAAGGCAATCTGTTG  259
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      181 CGCGGGCTGCCGGGACTCGGCGCGGGGACTGCATGGAGGGCCAAGGAGAGGAAGGCAATCTGTTG  240

Qy      260 GACGCGCAGGCCGGCGCAAATCCGGTCATACAACAGGGATCTCTGTTGGCCAGGAAGGGGCGGCTGG  319
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      241 GACGCGCAGGCCGGCGCAAATCCGGTCATACAACAGGGATCTCTGTTGGCCAGGAAGGGGCGGCTGG  300

Qy      320 ATTCCTCTGCCCCCGACCTGGCCTGGGGCTGGGGCGGCCATTGAGCTTGCGCCCAGGGGGCAAC  379
            |||||||||||||||||||||||||| |||| ||||| ||||||||||||||||||||||||||
Db      301 ATTCCTCTGCCCCCGACCTGGCCTGGACTTGCGAGGCCATTGAGCTTGCGCCCAGGGGGCAAC  360

Qy      380 CATCACTGCCATGCTCAGAAGGGTCCTGACAGTGCACTGTGACCCCAAGAAAGGGGAGGCC  439
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      361 CATCACTGCCATGCTCAGAAGGGTCCTGACAGTGCACTGTGACCCCAAGAAAGGGGAGGCC  420

Qy      440 CAGCGGCCAGCTGTATGTGAGCCTCTGCCATCTCCTGTTGTTGTTCATGATCGGGAGAAGTGGTT  499
            ||||| |||||||||||| |||||||||||||||| |||||||||||||||||||||||||| ||
Db      421 CACGCGCCAGCTGTATGTAGCCCTCTGCCATCTCGCCTGTTGTTCATGATCGGGAGAAGTGCTT  480

Qy      500 GGTGGGGTACCTGGCCACACAGCGTTGGGCTGTCATGAGCTGACTGAGCAGCCACACACCTGCACTGAC  559
            |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      481 GGTGGGGTACCTGGCCACACAGCGTTGGGCTGTCATGAGCTGACTGAGCAGCCACACACCTGCACTGAC  540
```

```
Qy  560  TTTG-CCAGCCATGCTCATCAG--CCTCTCTTCCCTCTCCCTTCTGATTGTCCTTCCC---GGCCAGCGCAC  615
         ||||  |||||||||||||||||   |||||| |||||||||||| |||||||||||||    ||||||| | ||
Db  541  TTTGCCCAGCCATGCTCATCAGCCCTCTTCCCTCTCCCTTCTGGATGTCCTTCCCCCGGCCAGCCACC        600

Qy  616  CAAGACCAATGAACTTTGG  633
         |||||||||||||  |||
Db  601  CAAGACCAATGAAACTTTG  618
```

RESULT 3
AU135737

```
AU135737
LOCUS       AU135737     783 bp    mRNA    linear    EST 02-AUG-2002
DEFINITION  AU135737 PLACE1 Homo sapiens cDNA clone PLACE1002782 5', mRNA
            sequence.
ACCESSION   AU135737
VERSION     AU135737.1   GI:10996276
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1 (bases 1 to 783)
  AUTHORS   Ota,T., Nishikawa,T., Suzuki,Y., Ishii,S., Saito,K., Kawai,Y.,
            Yamamoto,J., Wakamatsu,A., Nakamura,Y., Nagai,T., Sugano,S. and
            Isogai,T.
  TITLE     HRI human cDNA project
  JOURNAL   Unpublished (2000)
COMMENT     Contact: Takao Isogai
            Genomics Laboratory
            Helix Research Institute
            1532-3 Yana, Kisarazu, Chiba 292-0812, Japan
            Tel: 81-438-52-3975
            Fax: 81-438-52-3986
            Email: genomics@hri.co.jp.
            HRI human cDNA project; 5'- & 3'-end one pass sequencing: Helix
            Research Institute; cDNA library construction: Department of
            Virology, Institute of Medical Science, University of Tokyo, and
            Helix Research Institute.
FEATURES             Location/Qualifiers
     source          1..783
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="PLACE1002782"
                     /clone_lib="PLACE1"
                     /tissue_type="placenta"
                     /note="Vector: pME18SFL3"
BASE COUNT   157 a   245 c   217 g   155 t   9 others
ORIGIN

Query Match            32.2%;  Score 521;  DB 9;  Length 783;
Best Local Similarity  100.0%;  Pred. No. 1.9e-70;
Matches  521;  Conservative  0;  Mismatches  0;  Indels  0;  Gaps  0;

Qy  1046  GTGATCCTGGTGTTGATGGAAGGGACCCCCAAGGGCGTTGACTTCACACGTGTTCGTGAT  1105
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     1  GTGATCCTGGTGTTGATGGAAGGGACCCCCAAGGGCGTTGACTTCACACGTGTTCGTGAT    60

Qy  1106  CTGCTGTGCTGTCGGTGAGGGGGTGAGAACCAGCTGTGGGGCCACTGACG  1165
          |||||| ||||||||||| ||||||| |||||  | | |||||||||||||
Db    61  CTCGTGCTGTCGGTGGAGGGGGGTAGAACGCCTGCATATCTGGGCACTGACG   120

Qy  1166  GTGGGCCCAGCCTGTGTTCTGTCTGTGTCCACATCGCCATTGCTCAGAATACAGACGCCCAGGCT  1225
          |||||||||||||||  || ||| || |||||||||||||||||| |||||||||||||||||||
Db   121  GTGGGCCCAGCCTGTTCTGTGTCCACATCGCCATTGCTCGAAATACAGACGCCCAGGCT    180

Qy  1226  GTGCTGGAGAGAACAGCCAGCAGCCGCGCCCTCCAAGGGGAAGTTCCACTTCCACACGTGACCATC  1285
          ||||||||  ||||||||||||||| | |||||||||||||||||||||||||||||| ||||||||
Db   181  GTGCTGGAAGACAGCCAGCAGCCCCTCCAAGGGGAAGTTCCACTTCCACACCGTGACCATC    240

Qy  1286  CAGATCGAGGACTACTCGGAGGAGGACATGAAGAGCTGTCAGGGAGGGCCCCTCTGAC  1345
          ||||||||||||||||||||||||||||| |||||||||| |||||||||| ||||||
Db   241  CAGATCGAGGACTACTCGGAGGACCATGAAGGAGACTGTCAGGGAGGCCCCCTCAGAC    300
```

```
Qy  1346  TGACGTCGCTCAGCCCAGGCACCAAACTGGGGCATGGAACCATGGAGCTGCTGAAGGAGGAGGAACTGCAGGGTGGGCTGGACTGAG  1405
          |||| ||||||| ||||||  ||||||||| | |||   | |||| || |||| |||| ||| |||| ||| ||||| ||||||| |||
Db   301  TGACTGCTCACCCAGCCACCAAACTGGGGCATGAACAGGACCATGGCTGGCTGGTGGACTGAG    360
```

RESULT 4
BQ219887

```
BQ219887
LOCUS       BQ219887     931 bp    mRNA    linear    EST 02-MAY-2002
DEFINITION  AGENCOURT_7578202 NCI_CGAP_St1 Mus musculus cDNA clone
            IMAGE:6051584 5', mRNA sequence.
ACCESSION   BQ219887
VERSION     BQ219887.1   GI:20401287
KEYWORDS    EST.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1 (bases 1 to 931)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: Jeffrey E. Green, M.D.
            cDNA Library Preparation: Life Technologies, Inc.
            cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
            DNA Sequencing by: Agencourt Bioscience Corporation
            Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLAM13305 row: k column: 09
            High quality sequence stop: 761.
FEATURES             Location/Qualifiers
     source          1..931
                     /organism="Mus musculus"
                     /db_xref="taxon:10090"
                     /clone="IMAGE:6051584"
                     /clone_lib="NCI_CGAP_St1"
                     /lab_host="DH10B (T1-resistant)"
                     /note="Organ: stomach; Vector: pCMV-SPORT6; Site_1: SalI;
                     Site_2: NotI; Cloned unidirectionally. Primer: Oligo dT.
                     Average insert size 1.77 kb. Library constructed by Life
                     Technologies. Note: this is a NCI_CGAP Library."
BASE COUNT   181 a   271 c   258 g   219 t   2 others
ORIGIN

Query Match            31.6%;  Score 510.8;  DB 14;  Length 931;
Best Local Similarity  79.9%;  Pred. No. 6.5e-69;
Matches  634;  Conservative  0;  Mismatches  138;  Indels  21;  Gaps  2;

Qy   272  GCAATCCGGTCATACACGGGATCTCTGTGCGAGGAAGGGCTGGCTGGATTCCTCTGCC  331
          || | ||||| || ||   |||| || |||| |||| | ||||| | ||||   ||||
Db   116  GCCTCCCAGGTCCTTCTTAGGATCTTCTGTGGGAAGAGTGAAGCTAGCCGGGATCCCC----  170

Qy   332  CGACCTGGCCTGTCGACTTGCAGGCCATTGAGCTGGCTGCCCCAGAACGCAACCATCACTGCCAT  391
          |   |||  ||||| | ||||| | |||| ||||||||| ||||| |||||||||| | ||||||
Db   171  ------CTGTGGACCTGCCCGCGGCGTTGACCTGGCTGTCCAGAGCGAACCATTATTGCCAT  223

Qy   392  GCTCAGAAGGGTCCTGACAGTCACTGTGACCCCAAGAAAGGGAGAAGGCCCAGCGCCAGCTG  451
          || |||||||| | ||| |||||| | |||||||||| ||| || |||||| | ||| |||
Db   224  GCCCAGAAGGATTCTGGAAGTCACCCTGACCCCAGAGAAGCAGAGGGGCCCGACGGCAAACTC  283
```

RESULT 5

```
LOCUS       BE746716      828 bp      mRNA      linear   EST 15-SEP-2000
DEFINITION  601579148P1 NIH_MGC_9 Homo sapiens cDNA clone IMAGE:3928163 5',
            mRNA sequence.
ACCESSION   BE746716
VERSION     BE746716.1  GI:10160708
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 828)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: DCTD/DTP
            cDNA Library Preparation: Ling Hong/Rubin Laboratory
            cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
            DNA Sequencing by: Incyte Genomics, Inc.
            Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at: image.llnl.gov
            Plate: LLCM758  row: o  column: 12
            High quality sequence start: 3
            High quality sequence stop: 724.
FEATURES             Location/Qualifiers
     source          1..828
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:3928163"
                     /clone_lib="NIH_MGC_9"
                     /tissue_type="adenocarcinoma cell line"
                     /lab_host="DH10B (phage-resistant)"
                     /note="Organ: ovary; Vector: pOTB7; Site_1: XhoI; Site_2:
                     EcoRI; cDNA made by oligo-dT priming. Directionally
                     cloned into EcoRI/XhoI sites using the following 5'
                     adaptor: GGCACGAG(G). Size-selected >500bp for average
                     insert size 1.8kb. Library constructed by Ling Hong in
                     the laboratory of Gerald M. Rubin (University of
                     California, Berkeley) using ZAP-cDNA synthesis kit
                     (Stratagene) and Superscript II RT (Life Technologies)."
BASE COUNT      159 a      217 c      285 g      167 t
ORIGIN
```

```
Query Match     25.4%;  Score 410.8;  DB 12;  Length 828;
Best Local Similarity   74.3%;  Pred. No. 1.3e-53;
Matches   691;  Conservative   0;  Mismatches   72;  Indels   0;  Gaps   7;
```

```
Qy     32 CACGCGGGAGCCAGAGCCGGAGCCCGGAGCTTGCAGCAGTCACTGTGACCCCAAGAAAGGGGAAGGCC     91
          ||||||||||| | ||||||||||||||||||||||||||||||||||||||||||||||| ||||||
Db      1 CACGCGGGAGCCAGAGCCGGAGCCCGGAGCTTGCAGCAGTCACTGTGACCCCAAGAAAGGGGAAGGCC     60

Qy     92 AGGAGCCCGGCCGGAGCCGAGTGAGCCGCAGCCCTGCCGGG----------GCCCTCGGGGCAGGTGGG    142
          ||||||||||||||||||||||||||||||| ||||||||           ||||||||||||||||
Db     61 AGGAGCCGGGGCCGGAGCCGAGTGAGCCGCAGCCCTGCCGGG--------GCCCTCGGGGCAGGTGGG    120

Qy    143 TGAGCCGCCC-GGGACTCGGCGGCGGGGGGCGGCCGCAACTTTGAGGGGGCCCACTCGGCGG-GGCTCG    201
          | | ||||  |||||||||||||||||||| || |||||||||||||||||||||||||||  |||||
Db    121 TGCGCCACCGTGAGTCCCGCCGCGCAACTTTCAGGGGGCCCACTCGGCGG-GGCTCG    180

Qy    202 CGGCTGCC--GGGACTCGGCGGCGGGGGGCGGCCATGGA--GGCCAAGGAGGAGAAGCGGGGGGCTCG    259
          | || ||   ||||||||||||||||||||||| |||||   ||||||||||||||||||||||||||
Db    181 CGGCTGCCTCGGGGACTCGGCGGCGGGGGGCGGCCATGGATGGCCAAGGAGGAGAAGCGGGGGGCTCG    240

Qy    260 GACGGCCAGGCCGGCAAATCCGGTCATACACGGGATCTCTGTGGCAGGAGAGGGGGCTGGGTGG     319
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    241 GACGGCCAGGCCGGCAAATCCGGTCATACACGGGATCTCTGTGGCAGGAGAGGGGGCTGGGTGG     300

Qy    320 ATTCCTCTGCCCCGACCTGGACTTGCAGGCCATTGAGCGTGGCCTGGCCCAGAGCAAC     379
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    301 ATTCCTCTGCCCCGACCTGGACTTGCAGGCCATTGAGCGTGGCCTGGCCCAGAGCAAC     360

Qy    380 CATCACTGCCATCCTCAGAAGGGTCCTCGACAGTCACTGTGACCCCAAGAAAGGGGGTTCGTT     439
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    361 CATCACTGCCATCCTCAGAAGGGTCCTCGACAGTCACTGTGACCCCAAGAAAGGGGGTTCGTT     420

Qy    440 CACGCCCAGCTGTATGTAGCCCTGCCCATGTCGTTCATTGATCGGAGAAGTCGTT     499
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    421 CACGCCCAGCTGTATGTAGCCCTGCCCATGTCGTTCATTGATCGGAGAAGTCGT-     479

Qy    500 GGTGGGTGACCTGGCACACAGCTTGGCTGTCATGAGCTTGACGCAGCACACCTGCTCACTGAC     559
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    480 ----------------------------------------------------------------     479

Qy    560 TTTGCCAGCATGCTCATCAGCCTCTTCTCCCCTCCGGCCCAGCCACCAAG     619
          |||||||||||||||||||||||||||||||||||||||||||||||||
Db    480 ---------------------------------------------------     479

Qy    620 ACCATGAACTTGGCTGCCAGAGAGCTGAGAGCTGGGAGCCCTGGTGTCTCTGTACTGTCC     679
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    480 --------------------------------------------------------------     479

Qy    680 ATCTGGGTCGTGACGGGGGGTACTGGTGTACCTGGCCGCTGGAGGCCCCTGGGTCTCTGTACTGTGG     739
          |||||||||||||||||||||||||||||||  |||||||||||||||||||||||||||||||||
Db    514 ATCTGGGTCGTGACGGGGGCGGTACTGGTGTACCTGGCTGGTGGGAGGCCTGGGTCTGTGAACATC     573

Qy    740 TATGAAATTGACGGGGGGACCCCATGCTGATCACGTCGGGCGTGTGGCCTGTGAACATC     799
          |||||||||||
Db    ...
```

```
Qy    452 TATGTAGCCCTGCCCATGTCGTTCATTGATCGGAGAAGTCGTTGGTGGGTGACCTG     511
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    284 TATGTGGCCCTGCCCATCGTCGGTTGTTCATGATCATTCATTCATTTGGTGGGTACCTG     343

Qy    512 GCACAGCTTGGCTGTCATGAGACTGACGCAGCACACCTGCTCACTGACTGGAATCGCTCACTGACTTTGCCAGCATG     571
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    344 CTCATTAGCCTGTTGCCCCTGCTGGGTTCATGATCATTTTGGGTGGGTACCTG     403

Qy    572 CTCATCAGCCTCTTCTCCCCTCGGATGAGTGTCCCGGCCCAAGCCACCATGAAACTTT     631
          || ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    404 CTCATTAGCCTCTTCGCCCCTGTCCCGGTGTTCTCCGGACCAAGCCACCATGAACTTC     463

Qy    632 GGCTGGCGAGAGCTGAGATCTGGGAGCGGGCCCTGGTCTCTGTACTGTGTCCATCTGGGTGTG     691
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    464 GGCTGGCCGAGAGCCCAGAGTCCTTGTTGTCCGTGCTCTTCCATCTGGGTGGTG     523

Qy    692 ACGGGGGGTACTGGTGTACCTGGCGTGTGGACGGGGCGCATGCTCATCTCGGGGACTATGAAAATTGAC     751
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    524 ACTGGGGTGCTCATGTACCTACCGGTGATTTCTGGGGGTGTGGAGGATTATGAGAATCGAA     583

Qy    752 GGGGGGACCATGCTGATCACGTCGCGGCCGCGCCTGCTGGCCTGTGGCATCATCATTAATAATGGGGGGTTG     811
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    584 GGGGGACCATGTGGATCACTTCGGGTCGCTGCCTGCTGGCTGTGAACATCATTAATTATGGGGGTTG     643

Qy    812 ACCCTTCACCAGTCTCGGCCATGGGCACAGCCACGGCACCA------CCAACCAGCAG     862
          ||||||||||||||||||||||||||||||||||||||||||       ||||||||||
Db    644 GCCCTCCATCAGTCTGGACATGGGAAACACGCCGTGACGAGCAGCCAG     703

Qy    863 GAGGAGAAACCCCAGGCCGTCCGAGCGGCGCCTTCATCCATCCATGTGATAGCGGGACTTTATGCAGAGC     922
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    704 CAGCAGAAACCCCAGCGTCGGAGCTGCCTTCATTCATTCGATGTGATGCGGGACTTTGTGCAGAGC     763

Qy    923 ATGGGGTGTCCTAGTGGCCAGCCCTATATTTTATACTTCAAGCCAGAAATACAAAGTATGTAGAC     982
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    764 GTGGGGGTACTGGTGGCAGCCTATATCATATATACTTCAAGCCAGAAATACAAGTACGGGGAC     823

Qy    983 CCCATCTGCACCTTCGTCTTCTCCATCCTGGGGACAAACCTTGACAGTCCTGAGATCGAGA    1042
          |||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    824 CCCATCTGCACCTTCTCCTTGGTTGTCTTCNTTCATCCTGGTGTCTTGGGGACAATCCTGAGA     883

Qy   1043 GATGTGATCCTGG    1055
          ||||||||||||
Db    884 GACGTGATCTTGG     896
```

```
Db    574  TATGAAAATTGACGGGGGACCATGCTGATCACGTGGTCGGGCTGCGCTGTGGGCTGTGAACATC 633
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy    800  ATAATGGGGTTGACCCTTCACCAGTCTGGCCAGGCACGGGCACCGGACCACCAACCAG 859
           |||||||||||||||||   || || ||||||| || |||| ||| ||||
Db    634  ATAATGGGGTTGACCCTTCAGCAGTCT--------GGCATGGGCACGACGGGACAGTCAA 686

Qy    860  CAGGAGAGAGAACCCACGGTCCGGAGCTGCCTTCATCCATGTGATCGGCGACTTTATGCAG 919
           |||| |||| |  || |||||||||||||||||||| |   | | ||||| | ||||
Db    687  CACCAGGAGGAGAACCCAGGGTCGGAGCTGCTTCATCCATGTGAATCGGGGAAATTTAATGCA 746

Qy    920  AGCATGGGTGTCCTAGTGGGCACCCTTATATT 949
           ||||| ||||||||||||||||||| ||||||||
Db    747  GAGATTGGGTGT-CTAGTGGAAGGCCTATATT 775
```

```
RESULT 6
BF509598/c
LOCUS       BF509598                 394 bp    mRNA    linear    EST 06-DEC-2000
DEFINITION  UI-H-BI4-apf-e-01-0-UI.s1 NCI_CGAP_Sub8 Homo sapiens cDNA clone
            IMAGE:3087073 3', mRNA sequence.
ACCESSION   BF509598
VERSION     BF509598.1  GI:11592896
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 394)
  AUTHORS   NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
  TITLE     National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
            Tumor Gene Index
  JOURNAL   Unpublished (1997)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            The sequence contained an oligo-dT track that was present in the
            oligonucleotide that was used to prime the synthesis of first
            strand cDNA and therefore this may represent a bonafide poly A
            tail. cDNA Library Preparation: M.B. Soares Lab Clone distribution:
            NCI-CGAP clone distribution information can be found through the
            I.M.A.G.E. Consortium/LLNL at:
            www-bio.llnl.gov/bbrp/image/image.html
            Seq primer: M13 Forward
            POLYA=Yes.
FEATURES             Location/Qualifiers
     source          1..394
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:3087073"
                     /clone_lib="NCI_CGAP_Sub8"
                     /note="Vector: pT7T3D-Pac (Pharmacia) with a modified
                     polylinker; Site_1: Not I; Site_2: Eco RI; NCI_CGAP_Sub8
                     is a subtacted library derived from NCI_CGAP_Sub5. The
                     NCI_CGAP_Sub8 library had 2.5 million recombinants. A
                     single-stranded DNA preparation of NCI_CGAP_Sub5 was used
                     as a tracer in a subtractive hybridization with a driver
                     comprising: a pool of clones from NCI_CGAP_Sub5 (IMAGE
                     clone Ids 2732833-2737415, 3068040-3069191; 75% of the
                     driver population), a pool of clones from NCI_CGAP_Sub4
                     (IMAGE clone Ids 2723592-2729326; 25% of the driver
                     population), NCI_CGAP_Sub6 (pool AIF-AJU, IMAGE Ids
                     2728969-2733190; 25% of the driver population), and
                     NCI_CGAP_Sub7 (IMAGE Ids 3069192-3072238, 3081864-3084550
                     ; 25% of the driver population). Subtraction was
                     performed as previously described [Bonaldo, Lennon &
                     Soares (1996): Normalization and Subtraction: Two
                     Approaches To Facilitate Gene Discovery. Genome Research
                     6:791-806."
TAG_LIB=NCI_CGAP_Kid3
TAG_TISSUE=kidney
TAG_SEQ=AATGC"
```

```
RESULT 7
BE741089
LOCUS       BE741089                 662 bp    mRNA    linear    EST 15-SEP-2000
DEFINITION  601594088F1 NIH_MGC_9 Homo sapiens cDNA clone IMAGE:3948165 5',
            mRNA sequence.
ACCESSION   BE741089
VERSION     BE741089.1  GI:10155081
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 662)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: DCTD/DTP
            cDNA Library Preparation: Ling Hong/Rubin Laboratory
            cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
            DNA Sequencing by: Incyte Genomics, Inc.
            Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at: image.llnl.gov
            Plate: LLCM810  row: p column: 22
            High quality sequence stop: 648.
FEATURES             Location/Qualifiers
     source          1..662
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:3948165"
                     /clone_lib="NIH_MGC_9"
                     /tissue_type="adenocarcinoma cell line"
                     /lab_host="DH10B (phage-resistant)"
                     /note="Organ: ovary; Vector: pOTB7; Site_1: XhoI; Site_2:
                     EcoRI; cDNA made by oligo-dT priming. Directionally
                     cloned into EcoRI/XhoI sites using the following 5'
                     adaptor: GGCACGAG(G).  Size-selected >500bp for average
                     insert size 1.8kb. Library constructed by Ling Hong in
```

```
Query Match    24.1%;    Score 389.2;  DB 12;  Length 394;
Best Local Similarity  99.2%;   Pred. No. 3.4e-50;
Matches  391; Conservative    0;  Mismatches   3;   Indels    0;   Gaps    0;
```

```
Qy   1191  ACATGCCATTGCTCAGAAATACAGAACGACGCCAGGCGTGCTGCTGAAGACAGCCAGCAGCCGCC 1250
           |||| |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    394  ACATCCCCATTGCTCAGAAATACAGACGACGCCCAGGCGTGTCGTGAAGACAGCCAGCCGCC 335

Qy   1251  TCCAAGGGAAGTTCCACTTCCACCACCGTGACCATCCAGATCGAGGACTACTCGAGGACA 1310
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    334  TCCAAGGGAAGTTCCACTTCCACCACCGTGACCATCCAGATCGAGGACTACTCGGAGGACA 275

Qy   1311  TGAAGGACGTGTCAGGCATGCCAGGGCCCCTCAGACTGACTGCTCAGCCAGCCACCAACTG 1370
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    274  TGAAGGACGTGTCAGGCATGCCAGGGCCCCTCAGACTGACTGCTCAGCCAGCCACCAACTG 215

Qy   1371  GGGCATGAACAGAGACCTGCAGGTGGCTGGACTGAGTGTCCCCAGGCCACCAGGACTT 1430
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    214  GGGCATGAACAGAGACCTGCAGGTGGCTGGACTGAGTGTCCCCAGGCCACCAGGACTT 155

Qy   1431  TGCCTACCCCAGCTGTGTTATAAACCCAGTGTCCCCTCCCGCCCACTCCAGGA 1490
           ||||||||||||| |||||||||||||||||||||||||||||||||||||||
Db    154  TGCCTACCCCAGCTGTGTTGTTATAAACCCAGTGTCCCCTCCCGCCCACTCCAGGA 95

Qy   1491  ATGGAGCTCTTCCCCAGCCTCCCATCTGACTACAGCCAGGGTGGGGGACTCAGCGGGGTATA 1550
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     94  ATGGAGCTCTTCCCCAGCCTCCCATCTGACTACAGCCAGGGTGGGGGACTCAGCGGGGTATAA 35

Qy   1551  AGCTAGTGTGACCCTGCAAAAAAAAAAAAAAAAAAAAAA 1584
           ||||||||||||||||||||||||||||||||||||||
Db     34  AGCTAGTGTGACCCTGCAAAAAAAAAAAAAAAAAAAAAA 1
```

```
BASE COUNT      65 a     99 c     128 g     102 t
ORIGIN
```

RESULT 8

LOCUS   AU138346   598 bp   mRNA   linear   EST 02-AUG-2002
DEFINITION   AU138346 PLACE1 Homo sapiens cDNA clone PLACE1008371 5', mRNA
             sequence.
ACCESSION   AU138346
VERSION   AU138346.1   GI:10999867
KEYWORDS   EST.
SOURCE   human.
ORGANISM   Homo sapiens
             Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
             Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1 (bases 1 to 598)
AUTHORS   Ota,T., Nishikawa,T., Suzuki,Y., Ishii,S., Saito,K., Kawai,Y.,
             Yamamoto,J., Wakamatsu,A., Nakamura,Y., Nagai,T., Sugano,S. and
             Isogai,T.
TITLE   HRI human cDNA project
JOURNAL   Unpublished (2000)
COMMENT   Contact: Takao Isogai
             Genomics Laboratory
             Helix Research Institute
             1532-3 Yana, Kisarazu, Chiba 292-0812, Japan
             Tel: 81-438-52-3975
             Fax: 81-438-52-3986
             Email: genomics@hri.co.jp
             HRI human cDNA project; 5'- & 3'-end one pass sequencing; Helix
             Research Institute; cDNA library construction; Department of
             Virology, Institute of Medical Science, University of Tokyo, and
             Helix Research Institute.
FEATURES   Location/Qualifiers
     source   1..598
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="PLACE1008371"
                     /clone_lib="PLACE1"
                     /tissue_type="placenta"
                     /note="Vector: pME18SFL3"

BASE COUNT   123 a   200 c   158 g   114 t   3 others
ORIGIN

Qy   Query Match   20.1%;   Score 325.4; DB 9; Length 598;
      Best Local Similarity   99.7%; Pred. No. 1.6e-40;
      Matches   326; Conservative   0; Mismatches   1; Indels   0; Gaps   0;

Qy 1240 CAGCAGCCCGCCCTCCAAGGGAAGTTCCACTTCCACACCGTGACCATCCAGAATGGAGGACTA 1299
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1 CAGCAGCCCGCCCTCCAAGGGAAGTTCCACTTCCACACCGTGACCATCCAGAATGGAGGACTA   60

Qy 1300 CTCGGAGGACATGAAGGACACTGTCAGGCATGCCAGGGCCCCTCGACTGACTGCTCAGCCA 1359
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   61 CTCGGAGGACATGAAGGACACTGTCAGGCATGCCAGGGCCCCTCGACTGACTGCTCAGCCA 120

Qy 1360 GGCACCAACTGGGGCATGAAACAGGACCTGCACGGTGGCTGGACTGAGTGTCCCCCAGGCCC 1419
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   121 GGCACCAACTGGGGCATGAAACAGGACCTGCACGGTGGCTGGACTGAGTGTCCCCCAGGCCC 180

Qy 1420 AGCCAGGACTTTGCCTACCCCAGCTGTGTTATAAAACCAGGTCCCCGTCCTCGACCCTGCC 1479
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   181 AGCCAGGACTTTGCCTACCCCAGCTGTGTTATAAAACCAGGTCCCCGTCCTCGACCCTGCC 240

Qy 1480 CCACTCCAGGAAATGGAGAGTCTTCCCAGCCTCCAGTCTGACTACAGACCAGGGTGGGGACTC 1539
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||| |||
Db   241 CCACTCCAGGAAATGGAGAGTCTTCCCAGCCTCCAGTCTGACTACAGACCAGGGTGGGGACTC 300

Qy 1540 ACGGGGTATAAAGCTAGTGTGACCCTG 1566
          ||||||||||||||||||||||||||
Db   301 AGCGGGTATAAAGCTAGTGTGACCCTG 327

RESULT 9

LOCUS   AI925963   319 bp   mRNA   linear   EST 07-MAR-2000
DEFINITION   AI925963 wh12gl1.x1 NCI_CGAP_Kid11 Homo sapiens cDNA clone IMAGE:2380580 3',
             similar to TR:O62941 Q62941 ZINC TRANSPORTER ZNT-2. ;, mRNA
             sequence.
ACCESSION   AI925963
VERSION   AI925963.1   GI:5661927
KEYWORDS   EST.
SOURCE   human.

RESULT 8

ACCESSION   AU138346
DEFINITION   AU138346 PLACE1 Homo sapiens cDNA clone PLACE1008371 5', mRNA
             sequence.
LOCUS   AU138346   598 bp   mRNA   linear   EST 02-AUG-2002

BASE COUNT   116 a   185 c   241 g   120 t
ORIGIN

Query Match   23.1%;   Score 373.4; DB 12; Length 662;
Best Local Similarity   77.1%; Pred. No. 7.3e-48;
Matches   628; Conservative   0; Mismatches   31; Indels   156; Gaps   8;

Qy   32 CACGCGGAGCCCAGAGCCCGGAGCTGCCAGCC-GCACCGGGGAGCCGGGGGAGCCTCAGGGCCG   90
          ||||||| ||||||||||||||||||||||| |||||||||| |||||||||||||||||
Db   1 CACGCGGACCCAGAGCCCGGAGCTGCCAGCGCCAGCAGCCCTGGGGAGCTCAGGGCCG   60

Qy   91 CAGGAGCCCGGGCCGGAGGTCAGACCCGACCTTCGCGCGGCGGCCCTTGGGGTCGGGCC 150
          |||||| ||||||||||||||||||||||||| |||||| | ||
Db   61 CAGGAG-CGGGCCGGAGTGAGCGCA-CTCGCGGGGGCCCTTGGGGGCGGGTGGGT---AGCG 115

Qy   151 ACCGGGAGTCCCGCGGCGCAACTTTCAGGGCGCACTCGGGCGGGCCTGCGGGCTGCCG 210
          | || |||||||||||||||||||||||||||||||||||||| |||||||||
Db   116 CCACCGGAGTCCGCGCAACTTTCAGGGGCCCACTCGCGGGCGGGGGGG-TGCGGCGGTGCCCG 174

Qy   211 GGACTCGGCGGCGGCGGACTGCATGGAGGCCAAGGAGCACATCTGTTGTGACGGCCAGGCC 270
          | |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   175 GACTCGGCTGCGGGACTGCATGGAGGGCCAAGGAGAAGCACATCTGTTGTGGGACGGCCAGGCC 234

Qy   271 GGCAATCCGGTCATACACGGGGATCTCTGTGTGGGCCAGGAGAGAAGGGGGCTGCCTGCG 330
          | ||| ||||||||||||| ||||||||||||||||||||||||||||| ||||||
Db   235 GGCAATCCGGTCATACACGGGGATCTCTGTGTGGGCCAGGAAGGGGGGCTGCCTGCTGGGGGGT 294

Qy   331 CCGACCTGGCCTGGACTTGCAGGCCATTGAGCTTGGTGCTGCTGCCCAGAGCAGCAACCATCACTGCCA 390
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   295 CCGACCTGGCCTGGACTTGCAGGCCATTGAGCTGGCTGCTGCTGCCCAGAGCAGCAACCATCACTGCCA 354

Qy   391 TGCTCAGAGAGGTCTCTGACAGTCACTGTGACCCAAGAAGAAGGGGAAGGGGGCCCAGCGCCAGCT 450
          |||||||||||| |||||||||||||||||||||||||||| ||||||| |||||||||
Db   355 TGCTCAGAAGGGTCCTGACAGTCACTGTGACCCAAGAAGAAGGGCCCAGCGCCAGCT 414

Qy   451 GTATGTAGCCTCTCGCCATCTGCCTGTGTTGTTCATGATCGAGAAGTCGTTGGTGTGGGTACCT 510
          |||||||||||||||||||||||||| ||||||||||||||||||||||||||||||||
Db   415 GTATGTAGCCTCTGCCATCTGCCTGTGTTGTTCATGATCGGAGAAGTCGT--------- 462

Qy   511 GGCACACAGCTTGGCGTGCTGTCATGACTGACGCAGCACACCTGCTCACTGACTTTGCCAGCAT 570
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   463 --------------------TGAGATCTTGGGAGCGCCCTGGTCTCTTCCAATCTGGGTCGT 507

Qy   571 GCTCATCAGCCTCTTCTCCCTCTGGATGTCTCCCGGCCACCACCACAAGACCATGAACTT 630
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   508 GACGGGGTACTGGTGTCTACCTGGGCTTGTGGAGCGGGGGGTACTAGTGATGAAATTG 567

Qy   631 TGGCTGCAGAGAGGTGAGAGATCTTGGGAGCGCCCTGTCTCTGTACGTCCACTCTGGGTCGT 690
          ||||||||||| ||||||||||||||||||||||||||||||||||||||||||
Db   568 ACGGGGGGGACCAGCATTGCTGATCAGCGTCCGGCTCGGGGCTGTGTGAACATCATAAATGGGGG 627

Qy   691 GACGGGGGTACTGGTGTCTACCTGGC-TGTGGAGCGGGGGTGATCTCTGGGGACTATGAAATTG 749
          |||||||||||||||||||||||| |||||||||||||||||||||||||||||
Db   628 TTGACCCTTCACCAGTCTGGCCATGGGCACACCA 662

Qy   750 ACGGGGGGGACCATGCTGATCACGTCGGGCTGCTGCCTGTGAGCGGGGGGGGGACTCATAAT-GGGG 808
          |||||||| ||||||||||||||||||||||||||||||||||||||||||||||||
Db   568 ACGGGGGGGACCATGCTGATCACGTCGGGCTCGGTGCTGTGTGAACATCATAAATGGGGG 627

Qy   809 TTGACCCTTCACCAGTCTGGCCATGGGCACAGCCA 843
          ||||||||||||||||||||||||||||||||||
Db   628 TTGACCCTTCACCAGTCTGGCCATGGGCACACCA 662

RESULT 10

BB620842

LOCUS   BB620842   599 bp   mRNA   linear   EST 26-OCT-2001
DEFINITION   BB620842 RIKEN full-length enriched, 13 days embryo male testis Mus
             musculus cDNA clone 603043C21 5', mRNA sequence.
ACCESSION   BB620842
VERSION   BB620842.1 GI:16459747
KEYWORDS   EST.
SOURCE   Mus musculus
  ORGANISM   Mus musculus
             house mouse.
             Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
             Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1 (bases 1 to 599)
  AUTHORS   Arakawa,T., Carninci,P., Fukuda,S., Furuno,M., Hanagaki,T., Hara,A.,
             Hiramoto,K., Hori,F., Ishii,Y., Ito,M., Kawai,J., Konno,H., Kouda
             ,M., Koya,S., Matsuyama,T., Miyazaki,A., Nomura,K., Ohno,M.,
             Okazaki,Y., Okido,T., Saito,R., Sakai,C., Sakai,K., Sano,H., Sasaki
             ,D., Shibata,K., Shinagawa,A., Shiraki,T., Sogabe,Y., Suzuki,H.,
             Tagami,M., Tagawa,A., Takahashi,F., Takeda,Y., Tanaka,T., Toya,T.,
             Muramatsu,M. and Hayashizaki,Y.
  TITLE   RIKEN Mouse ESTs (Arakawa,T., et al. 2001)
  JOURNAL   Unpublished (2001)
COMMENT   Contact: Yoshihide Hayashizaki
             Laboratory for Genome Exploration Research Group, RIKEN Genomic
             Sciences Center(GSC), Yokohama Institute
             The Institute of Physical and Chemical Research (RIKEN)
             1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan
             Tel: 81-45-503-9222
             Fax: 81-45-503-9216
             Email: genome-res@gsc.riken.go.jp
             URL:http://genome.gsc.riken.go.jp/
             Carninci,P., Shibata,Y., Hayatsu,N., Sugahara,Y., Shibata,K., Itoh
             ,M., Konno,H., Okazaki,Y., Muramatsu,M. and Hayashizaki,Y.
             Normalization and subtraction of cap-trapper-selected cDNAs to
             prepare full-length cDNA libraries for rapid discovery of new
             genes. Genome Res. 10 (10), 1617-1630 (2000)
             wagi,K., Fujiwake,S., Inoue,K., Togawa,Y., Izawa,M., Ohara,E.,
             Watahiki,M., Yoneda,Y., Ishikawa,T., Ozawa,K., Tanaka,T., Matsuura
             ,S., Kawai,J., Okazaki,Y., Muramatsu,M., Inoue,Y., Kira,A. and
             Hayashizaki,Y.
             RIKEN integrated sequence analysis (RISA) system-384-format
             sequencing pipeline with 384 multicapillary sequencer. Genome Res.
             10 (11), 1757-1771 (2000)
             Konno,H., Fukunishi,Y., Shibata,K., Itoh,M., Carninci,P., Sugahara
             ,Y. and Hayashizaki,Y.
             Computer-based methods for the mouse full-length cDNA
             encyclopedia: real-time sequence clustering for construction of a
             nonredundant cDNA library. Genome Res. 11 (2), 281-289 (2001)
             Kondo,S., Shinagawa,A., Saito,T., Kiyosawa,H., Yamanaka,I., Aizawa
             ,K., Fukuda,S., Hara,A., Itoh,M., Kawai,J., Shibata,K. and
             Hayashizaki,Y.
             Computational Analysis of Full-Length Mouse cDNAs Compared with
             Human Genome Sequences. Mamm. Genome. 12, 673-677 (2001)
             Please visit our web site (http://genome.gsc.riken.go.jp) for
             further details.
FEATURES   Location/Qualifiers
     source   1..599
                     /organism="Mus musculus"
                     /strain="C57BL/6J"
                     /db_xref="taxon:10090"
                     /clone="603043C21"
                     /clone_lib="RIKEN full-length enriched, 13 days embryo
                     male testis"
                     /sex="male"
                     /tissue_type="testis"
                     /dev_stage="13 days embryo"
                     /lab_host="DH10B"
                     /note="Site_1: SalI; Site_2: BamHI; cDNA library was
                     prepared and sequenced in Mouse Genome Encyclopedia
                     Project of Genome Exploration Research Group in Riken
                     Genomic Sciences Center and Genome Science Laboratory in
                     RIKEN. Division of Experimental Animal Research in Riken
                     contributed to prepare mouse tissues. 1st strand cDNA was
                     primed with a primer [5'-
                     GAGAGAGAGAGATCCAAGAGCTCTTTTTTTTTTTTTTTTTTTTVN 3'], cDNA was
                     prepared by using trehalose thermo-activated reverse

---

ORGANISM   Homo sapiens
             Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
             Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1 (bases 1 to 319)
  AUTHORS   NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
  TITLE   National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
             Tumor Gene Index
  JOURNAL   Unpublished (1997)
COMMENT   Contact: Robert Strausberg, Ph.D.
             Email: cgapbs-r@mail.nih.gov
             Tissue Procurement: Christopher Moskaluk, M.D., Ph.D., Michael R.
             Emmert-Buck, M.D., Ph.D.
             cDNA Library Preparation: M. Bento Soares, Ph.D.
             cDNA Library Arrayed by: Greg Lennon, Ph.D.
             DNA Sequencing by: Washington University Genome Sequencing Center
             Clone distribution: NCI-CGAP clone distribution information can be
             found through the I.M.A.G.E. Consortium/LLNL at:
             www.bio.llnl.gov/bbrp/image/image.html
             Seq primer: ~40UP from Gibco.
             Insert Length: 395   Std Error: 0.00
FEATURES   Location/Qualifiers
     source   1..319
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:2380580"
                     /clone_lib="NCI_CGAP_Kid11"
                     /lab_host="DH10B"
                     /note="Organ: kidney; Vector: pT7T3D-Pac (Pharmacia) with
                     a modified polylinker; Site_1: Not I; Site_2: Eco RI;
                     Plasmid DNA from the normalized library NCI_CGAP_Kid3 was
                     prepared, and ss circles were made in vitro. Following HAP
                     purification, this DNA was used as tracer in a subtractive
                     hybridization reaction. The driver was PCR-amplified cDNAs
                     from a pool of 5,000 clones made from the same library
                     (cloneIDs 1322376-1323911, 1456007-1456775, and
                     1500052-1502855). Subtraction by Bento Soares and M.
                     Fatima Bonaldo."

BASE COUNT   56 a   81 c   95 g   87 t
ORIGIN

Query Match   18.4%;   Score 298.2;   DB 9;   Length 319;
Best Local Similarity   95.9%;   Pred. No. 2.9e-36;
Matches   306;   Conservative   0;   Mismatches   13;   Indels   0;   Gaps   0;

Qy 1273 CACCGTGACCATCCAGATCGGAGGACTACTCGGAGGACATGAAGGAGGACTGTCAGGCATCCA 1332
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   319 CACCGTGACCATCCAGATCGGAGGACTACTCGGAGGACATGAAGGAGGACTGTCAGGCATCCA 260

Qy 1333 GGGCCCCTCAGACTGACTGCTCTCAGCCAGGCACCACAACTCGGGCGCATGAACAGGACCTCGCAGG 1392
         ||||||||||||| |||| || ||||||||| |||||||||||||||||||||||||||||||||
Db   259 GGGCCCCTCAGAGTGATTGGTTCAGCCAGGCCCCAACTCGGGCGCATGAACAGGACCTCGCAGG 200

Qy 1393 TGGCTGGACTGAGTGGTCCCCCAGGCCCCAGGGACTTTGCCTACCCCAGCTGTGTTATA 1452
         ||| |||| ||||||| |||||||||||||||||||||||||||||||||||||||||||
Db   199 TGGGTGGGACTGAGTGTCCCCCAGGCCCCAGGGACTTTGCCTACCCCAGCTGTGTTATA 140

Qy 1453 AACCAGTCCCCCTCCTGACCCTCTGCCCCACTCCAGGAATGGAGCTCTTCCCAGCCTCCC 1512
         ||||||| |||||||||||| || |||||||||||||||||||||| ||||||||||||
Db   139 AACCAGGTCCCCCTCCTGACCTTGCCCCACTCCAGGAATGGAGCTTTTCCCAGCCTCCC 80

Qy 1513 ATCTGACTACAGCCAGGGGACTGGGGACTCAGCGCGGGTATAAAAGCTAGTGTGACCCTGAAAAA 1572
         || |||||||| ||||| | ||||||||||||| ||||||||||| ||| ||| ||| |||||
Db   79 ATTTGACTACACCCAGGGTGGGGACTCAGCCGCGGGTATAAAACCCGTGTGGCCCTTAAAAA 20

Qy 1573 AAAAAAAAAAAAAAAAAA 1591
         ||||||||||||||||||
Db   19 AAAAAAAAAAAAAAAAAA 1

TITLE   RIKEN Mouse ESTs (Arakawa,T., et al. 2001)
             Muramatsu,M. and Hayashizaki,Y.
             Tagami,M., Tagawa,A., Takahashi,F., Takeda,Y., Tanaka,T., Toya,T.,
             D., Shibata,K., Shinagawa,A., Shiraki,T., Sogabe,Y., Suzuki,H.,
             Okazaki,Y., Okido,T., Saito,R., Sakai,C., Sakai,K., Sano,H., Sasaki
             M., Koya,S., Matsuyama,T., Miyazaki,A., Nomura,K., Ohno,M.,
             Hiramoto,K., Hori,F., Fukuda,S., Furuno,M., Hanagaki,T., Hara,A.
AUTHORS   Arakawa,T., Carninci,P., Fukuda,S., Furuno,M., Hanagaki,T., Hara,A.
             1 (bases 1 to 567)
REFERENCE   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
             Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
ORGANISM   Mus musculus
SOURCE   house mouse.
KEYWORDS   EST.
VERSION   BB649221.1 GI:1648476
ACCESSION   BB649221
             musculus cDNA clone C13006214 5', mRNA sequence.
DEFINITION   BB649221 RIKEN full-length enriched, 16 days embryo head Mus
LOCUS   BB649221   567 bp   mRNA   linear   EST 26-OCT-2001

RESULT 11

Db   546 CAAAGGGGGCACCCATGGGTGGTCACTTCGGGCTGCCCTTGGGGTGGGAA 595
         |||||||||||||||||||||||||||||||||||||||||||||||||
Qy   748 TGACGGGGG--GACCATGCTGATCACGTCGGGGCTGCGCTGTGCCTTGAA 795

Db   486 GGTGACTGGGGTGGTCGTGACTACCTGCGCAACCGTTGAATTTCTGGGGA 545
         ||||| |||| ||||||||||||||||||||||||||||| |||||
Qy   688 CGTGACGGGGGTACTGGGTGGTACCTGGCGTGTGGGAGCGGGGACTATAGAAAT 747

Db   426 TCGGGCTGCCATCGAGCCGAGAGTCCTTGGGAGGCCCTTGGTTGGGTCCCGCTTCATTGGGT 485
         | ||||||||| ||||||||||||||||||||| ||||| |||
Qy   630 TTGGCTGGCAGAGAGCTGAGAGATCTTGGGGAG--CCCTGGTGTCTGTACTGTCCATCTGGGT 687

Db   366 TGCTCATTAGCCTCTTCTCGCCCCTTCTGGGTGTGTCTTTTTCCCACCCACCACCATGAACT 425
         ||||||| |||||||||||||||||||||||||||||||||||||||
Qy   570 TGCTCATCACGCCTCTTCTCTCCTCTTGATTGTCTCCCGGGCCCACCACCACAGACCGAACT 629

Db   306 TGGACAGAGGCCTGGCCATCATGACCACAGAGCGCTGCTCACGGGAGGATTTTGCCAGCA 365
         ||||| |||||||| |||||||||||||||||||||||| |||||
Qy   510 TGGCACACAGCGTTGGCGTGTCATGGACTGACGCAGCACACCTCTCACTGACTTTGCCAGCA 569

Db   246 TCTATGTGGGCTCTGCCATCTGCCCTCGGTGTTCATGAGTGATTGGGAGCAAAATCATTGGTGGGTACC 305
         | |||| |||||| || ||| ||||| ||||||||||||||||||||||||||||||||||||
Qy   450 TGTATGTAGCCTCTGCCACTCGCCTGTTGTTCATGAATCGGAGAAGTCGTTGGTGGGTACC 509

Db   186 ATGCCCAAGAGGATTCTGGAAGTACACCCTGACCCCGAGAAGGCAGAGGGCCCACGCAAAC 245
         ||||||||||||||||||||||||| |||||| |||||| |||||||||||||||||||||
Qy   390 ATGCTCAGAAGGGTCCTCGACAGTCACTGTTGACCCCAAGAAGAAGGGGAAGGGGAAGGGCGGCCAGC 449

Db   135 ------CTGTGAGACCTGCCCGCGGCCGGTTGAGCTGGTCGTCCAGAGCCAACCAATTATTGCC 185
         ------| ||||| ||||||||||||||| ||| |||||||||||
Qy   330 CCCGACCTGGCCTGGACTTGCAGGCCATTGAGCTGGCTGGCTGCCCAGAGCAAACCATCACTGCC 389

Db   78 CTGCCTGCCCTGTCCTTCTTAGGATCTCTGTGGAAGAGTGAAGCTAGCCGGATCCCTC--- 134
         || ||| |||| ||| ||||||||||| ||| ||| ||||||| ||||||||| ||||
Qy   270 CGGCAATCCGGTCATACACGGAGGATCCTCGTGTGGCAGGAAGGGGCTGGCTGGGATTCCTGC 329

BASE COUNT   111 a   173 c   177 g   138 t

ORIGIN

Qy   173 c   177 g   138 t

Matches 405; Conservative   0; Mismatches 109; Indels 16; Gaps   3;

Best Local Similarity 76.4%;   Pred. No. 1.2e-35;
Query Match   18.28;   Score 293.6; DB 10; Length 599;

FEATURES   Location/Qualifiers
             source   1..567
                             /organism="Mus musculus"
                             /strain="C57BL/6J"
                             /db_xref="taxon:10090"
                             /clone="C13006214"
                             /clone_lib="RIKEN full-length enriched, 16 days embryo
                             head"
                             /sex="mixed"
                             /tissue_type="head"
                             /dev_stage="16 days embryo"
                             /lab_host="DH10B"
                             /note="Site 1: SalI; Site 2: BamHI; cDNA library was
                             prepared and sequenced in Mouse Genome Encyclopedia
                             Project of Genome Exploration Research Group in Riken
                             Genomic Sciences Center and Genome Science Laboratory in
                             RIKEN. Division of Experimental Animal Research in Riken
                             contributed to prepare mouse tissues. 1st strand cDNA was
                             primed with a primer (5'
                             GAGAGAGAAGGATCCAAGAGCTCTTTTTTTTTTTTTTTTTTVN 3'), cDNA was
                             prepared by using trehalose thermo-activated reverse
                             transcriptase and subsequently enriched for full-length by
                             cap-trapper. Second strand cDNA was prepared with the
                             primer adapter of sequence (5'
                             GAGAGAGGATCCGAGTTAAATTAAATTAATCCCCCCCCCCCC 3'). cDNA
                             was cloned into the XhoI and BamHI sites. Vector: a
                             modified pBluescript KS(+) after bulk excision from Lambda
                             FLC I"

BASE COUNT   98 a   177 c   161 g   128 t   3 others

ORIGIN

Query Match   17.28;   Score 278.4; DB 10; Length 567;
Best Local Similarity 77.7%;   Pred. No. 2.5e-33;
Matches 352; Conservative   0; Mismatches 89; Indels 12; Gaps   1;

Qy   272 GCAATCCGGTCATACACGGATCCTCGTGTGGCAGGAAGGGGCTGGCTGGGATTCCTGCTGCC 331
         || ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   126 GCCTCCCAGGTCCTTCTTAGGATCTCTGTGGAAGAGTGAAGCTAGCCGGATCCCTC---- 180

RESULT 12

```
Qy    332 CGACCTGGCCTGGACTTGCAGGCGCCATTGAGCTGGCTGCGCCCAGAGCAACCATCACTGCCAT 391
          |||||| ||| |||||||||||||||||||||||||||||| |||||||||||| ||||||
Db    181 ------- CTGTGGACCTGGCCGCCGGTTGAGCTGGCTGCTCCAGAGCAACCATTATTGCCAT 233

Qy    392 GCTCAGAAGGGTTCCTGACAGTCACTGTTGACCCCAAGAGAGGGAAGGCCCAGCCCCAGCTG 451
          |||||||||||||||||||||| ||||||||||||||||||||||||||||||||||||||
Db    234 GCCCAGAAGGGAATTCTGGAAGTCACCGTTGACCCCGAGAGAAGCACAGGGCCCGACGCAAACTC 293

Qy    452 TATGTGAGCCTCTTGCCATCTGCCTGTTGTTCATGATCGGAAGTCGTTGGTGGGGTACCTG 511
          |||||||||||||||||||||| ||||||||||||| |||||||||||||||||||||||
Db    294 TATGTGGGCCTCTCCGCCATCTGCCTGTTGTTCATGATCATTGGTGGTGGTGGGGTACCTG 353

Qy    512 GCACACAGCTTGGCTGTCATGACGCACGCACACCTGCTCACTGACTTTGCCAGCATG 571
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    354 GCACAGAGCCTGGCCATCATGACGCACGCTGCCACCTGCTCACGGATTTGCCAGCATG 413

Qy    572 CTCATCAGCCCTCTTCTCCCCGGATGTCCTCCCGGGCCCAGCCACCAAAGACCATGAACTT 631
          |||||||||||||||||||||||||||||| |||||| |||||||||||||||||||||
Db    414 CTCATTAGCCCTCTTCGCCCCTCCGGGTTCTTCCGACCAGCCACCATGAACCATCAGTTA 473

Qy    632 GGCTGGCAGAGAGCTGAAGATCTTGGGAGACCCCTGGCTGTCTCTGTCCATCTGGGTGCTG 691
          |||||||||||||||||||||||||||| |||| ||||||||||||||||||||||
Db    474 GGCTGGCCATCGAGCCGAGATACCCTTGGAGGACCTTGTTGTTCCCGGGCCTTTCCATCTGGGTGGGG 533

Qy    692 ACGGGGGGTACTGGTGGTCATTGTTACCTGGCTGTGTGGAGCGG 724
          |||||||||||||||||| ||||||||||||||||||||||
Db    534 ACTGCGGGNGCCNGNGNGTACCTGGCCGCGGGCAGCCGG 566
```

LOCUS   AL651875
DEFINITION AL651875 XGC-gastrula Silurana tropicalis cDNA clone TGas037a03 5',
            mRNA sequence.
ACCESSION   AL651875
VERSION   AL651875.1 GI:17662169
KEYWORDS   EST.
SOURCE   western clawed frog.
  ORGANISM Silurana tropicalis
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Amphibia; Batrachia; Anura; Mesobatrachia; Pipoidea; Pipidae;
            Xenopodinae; Silurana.
REFERENCE   1 (bases 1 to 674)
  AUTHORS   Huckle,E., Taylor,R., Ashurst,J.L., Zorn,A.M. and Rogers,J.
  TITLE   Sanger Xenopus tropicalis EST project 2001
  JOURNAL   Unpublished (2001)
COMMENT   Contact: Huckle E
            Sanger Centre
            Hinxton, Cambridgeshire, CB10 1SA, UK
            Email: trop@sanger.ac.uk
            Sanger Xenopus tropicalis EST project 2001
            TROPICALIS_SEQUENCE_ID: TGas037a03.sp6
            Sequencing primer: SP6
            This sequence is from a Xenopus Gene Collection (XGC) library
            constructed by Aaron M. Zorn.
FEATURES   Location/Qualifiers
     source   1..674
                     /organism="Silurana tropicalis"
                     /db_xref="taxon:8364"
                     /clone="TGas037a03"
                     /clone_lib="XGC-gastrula"
                     /dev_stage="gastrula (stages 10.5-13 mixed)"
                     /lab_host="Escherichia coli XL1-blue"
                     /note="Vector: pCS107; Site_1: EcoRI; Site_2: NotI; cDNA
                     was oligo dT primed from 5ug of poly A+ RNA from stages
                     10-13 gastrulae. EcoRI-NotI cut cDNA was then ligated
                     into pCS107 with EcoRI at the 5' end and NotI at the 3'
                     end."
BASE COUNT   175 a   147 c   187 g   164 t   1 others
ORIGIN

Query Match   16.8%; Score 271.8; DB 9; Length 674;

Best Local Similarity 72.6%; Conservative 0; Mismatches 137; Indels 1; Gaps 1;
Matches 365;

RESULT 13

```
Qy    344 GACTTGCAGGCCCATTGAGCTGGCTGCGCCCAGAGCAACCATCACTGCCATCCTCAGAAGGGT 403
          || ||||||||||||||||||||||||||||||||| |||||||||||||||| |||||
Db    140 GATCTCAATGGTATTGAACTGGGTGCTCATCAGGGGGAACCAGCCACTGTTCACTCT-AACCGGGA 198

Qy    404 CCTGACAGTCACTGTTGACCCCAAGAGAGGGAAGGCCCAGCCCCAGCTGTATGTGAGCCTC 463
          |||            |||||||||||||||||||||||||||||||||||||||||||
Db    199 ACTGGATGACTGGAAATGCCAAAGAGAGAAAAAAAAGGCTCGTCGAAAAGCTGTATGTGGGCATCA 258

Qy    464 GCCATCTGCCTGTTGTTCATGATCGGAAGTCGTTGGTGGGGTACCTGGCACACAGCTTG 523
          ||||||||||||||||||||||||| || |||||||||||||||||||||||||||
Db    259 GCGGTCTTGCCTGGTGTTCATGATCGGAGGGGGTACCTGGCACAGAGCCTG 318

Qy    524 GCTGTCATGACGCACGCACACCTGCTCACTGACTTTGCCAGCATGCTCATCAGCCCTC 583
          ||||||||||||||||||||||||||||| ||||||||||||||||||||||| |||
Db    319 GCAATCATGACACAGAGCTGCCACCTTCTGACAGATTTTGCTACTGATGATCAGTTTA 378

Qy    584 TTCTCCCCTCTGGATGTCCTCCCGGGCCCAGCCACCAAAGACCATGAACTTGGCTGGCAGAGA 643
          |||| ||||||||||||| |||||||||||||||||||||||||||||||||||||||
Db    379 TTTGCTCTGTGGATGTCCTCCAGCCAGCCACTGACTACTGACTTTGGCTGGCCATCGA 438

Qy    644 GCTGAGATCTTGGGAGACCCCTGGCTGTCTCTGTACTGTGTCCATCTGGGTGCTGGCGAGAGA 703
          |||||||| ||||||||||||||||||||| |||||||||||||||||||||||||||
Db    439 GCAGAAATTTTGGGAGGACGGGCTTCCTCTGTCCGTCTGAACTGGGTTGTGTGACTGGAGTCTTG 498

Qy    704 GTGTACCTCGGCTGTGGAGGGGGGTTGATCTCTGGAGGACTATGAAAATTGACCGGGGACCATG 763
          |||||| |||||||||||||||||||| |||||||||||||||||||||||||||||||
Db    499 GTCTACCTAGCCTGTGCAGGCGGGCTATAATATCATTCAGGAGGATTACGGAGATTGAGGGCTATG 558

Qy    764 CTGATCACGTCGGCGGCTGCCTGTGCTGGTGAACATCATAATGGGGGTTGACCCTTCACCAG 823
          ||||||||||||||||||||||||||||||||||||||||||||||| |||||||||||
Db    559 CTGATCACCTCGGCGCCTGTCCCAGTTGCAGTCAAATATCATAATAATGGGGGTAACTCTTCATCAG 618

Qy    824 TCTGCCCATGGGCACAGCCACACGG 846
          |||||||||||||||||||||||
Db    619 ACAGGGCATGGACACAGTCACGG 641
```

LOCUS   BE666434   524 bp   mRNA   linear   EST 25-APR-2001
DEFINITION 149858 MARC 4BOV Bos taurus cDNA 5', mRNA sequence.
ACCESSION   BE666434
VERSION   BE666434.1 GI:10027025
KEYWORDS   EST.
SOURCE   cow.
  ORGANISM Bos taurus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovoidea;
            Bovidae; Bovinae; Bos.
REFERENCE   1 (bases 1 to 524)
  AUTHORS   Smith,T.P.L., Grosse,W.M., Freking,B.A., Roberts,A.J., Stone,R.T.,
            Casas,E., Wray,J.E., White,J., Cho,J., Fahrenkrug,S.C., Bennett
            G.L., Heaton,M.P., Laegreid,W.W., Rohrer,G.A., Chitko-McKown,C.G.,
            Pertea,G., Holt,I., Karamycheva,S., Liang,F., Quackenbush,J. and
            Keele,J.W.
  TITLE   Sequence evaluation of four pooled-tissue normalized bovine cDNA
            libraries and construction of a gene index for cattle
  JOURNAL   Genome Res. 11 (4), 626-630 (2001)
  MEDLINE   21180013
COMMENT   Contact: Smith TPL
            USDA, ARS, US Meat Animal Research Center
            PO Box 166, Clay Center, NE 68933-0166, USA
            Tel: 402 762 4366
            Fax: 402 762 4390
            Email: smith@email.marc.usda.gov
            Single pass sequencing. Bases called and alt_trimmed with phred
            v0.980904.e. Vector identified by cross_match with the -minscore 18
            and -minmatch 12 options.
            PCR PRimers
            FORWARD: AGGAAACAGCTATGACCAT

BACKWARD: GTTTTCCAGTCACGACG
Plate: 58   row: I   column: 16
Seq primer: ATTTAGGTGACACTATAG.

FEATURES   Location/Qualifiers
     source   1..524
                  /organism="Bos taurus"
                  /db_xref="taxon:9913"
                  /clone_lib="MARC 4B0V"
                  /tissue_type="pooled"
                  /lab_host="DH10B"
                  /note="Vector: pCMV SPORT6; Site_1: NotI; Site_2: SalI;
                  Library made from pooled tissue from day 20 and day 40
                  embryos."

BASE COUNT   97 a   172 c   155 g   100 t
ORIGIN

Qy   1066 AGGGACCCCCAAGGGACGGTGTGACTTTACACGCCTGTGCAGAAATCTGCTTGCTGCTTCGGGTGGAGGG 1125
          ||||||||||||||||||||||||||||||||||||||| ||||||||||||||||||||||||||||
Db   2 AAGGGACCCCCAGGGGTGTGAGACTTTACAGGCCGTGCAGAAATCTGCTTGCTGTGTCCGGTGAGG 61

Qy   1126 GGTAGAAGCCCTGCACAGCCTGCATATCTGGGCACTGACGGTGGGCCCAGCCGTGTTCTGTC 1185
          ||||||||| ||||||||||||||||||||||||| ||||||||||||||||| ||||||||
Db   62 CGTGGGAAGCCTTGCACAGCCTGCACATCTGGGCACTCGGCCCTGGCCCAGCCTATACTGTC 121

Qy   1186 TGTCCACATCGCCATTGCTCAGAAATACAGACGCGCCTGTGCTGCTGAAGACGCCAGCAG 1245
          |||||||||||||||||||||||||||||||||||||||||| ||| ||||||| ||||
Db   122 TGTCCACATCGCCATCGCTCAGAAATCGAGATNCGAGTGCTGCTGAAGGCGAGCGA 181

Qy   1246 CCGGCTCCAAGGGAAGTTCACTTCCACACCGTGACCATCCAGATCGAGAGACTACTCGGA 1305
          |||||||||||||| |||||||||||||||||||||||||||||||||||||| |||||
Db   182 CCGCCTGCAGGAGATGTTCCATTTCCACACCATGGCCATCCAGATCGAGGACTACTCTGA 241

Qy   1306 GGACATGAAGGACCTGTCAGGCCATGCCAGGGCCCCTCAGCACTGACTCACCCAGGCACC 1365
          ||||||||||||||||||||||||||||||||||||||||| | ||||||| |||||||
Db   242 GGACATGAAGGACTGTCAGTCATGCGGCGGCCCCCTCGGCTCTGACCACCAGGCATC 301

Qy   1366 AACTGGGGCATGAACAGGACCTGCAGGTGGCTGGACTGAGTGT---------CCCCC 1413
          |||| || |||||| |||| | ||| |||||||||||| |||| |||||
Db   302 AACGGGGTGGGGA--CAGGCCTGCCGATGGCTGGATCGAGTGTCCCCCACCCCACCCCC 359

Qy   1414 AGGCCCAGCCAGGACGACTTTGCCTACCCCACCCCAGCTGTGTTATAAACCAGGTCCCCTCCTGACC 1473
          |||||||||||| |||||||||||||| || ||| |||||| |||||||||| || |||||
Db   360 AGGCTCAGGCAGGTCTCTGCCTGCCCTGGGTTATGAACCAGGTCCTTCTCTTCCTGACC 419

Qy   1474 TCTGCCCCA 1482
          ||| |||||
Db   420 GCTGCTCCA 428

RESULT 14
BQ384277
LOCUS   BQ384277   608 bp   mRNA   linear   EST 22-MAY-2002
DEFINITION NISC_mn05g11.y1 NICHD XGC Ovl Xenopus laevis cDNA clone
            IMAGE:5049404 5', mRNA sequence.
ACCESSION   BQ384277
VERSION   BQ384277.1 GI:21071964
KEYWORDS   EST.
SOURCE   African clawed frog.
  ORGANISM Xenopus laevis
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Amphibia; Batrachia; Anura; Mesobatrachia; Pipoidea; Pipidae;
            Xenopodinae; Xenopus.
REFERENCE   1 (bases 1 to 608)
AUTHORS   NIH-XCG http://image.llnl.gov/image/html/xenopuslib_info.shtml.
TITLE   National Institute of Child Health and Human Development, National
            Cancer Institute, Xenopus Gene Collection
JOURNAL   Unpublished (2002)
COMMENT   Contact: Robert Strausberg, Ph.D.
            Email: cgapbs.r@mail.nih.gov

cDNA Library Preparation:
cDNA Library Arrayed by: The I.M.A.G.E. Consortium/LLNL
DNA Sequencing by: National Institutes of Health Intramural
Sequencing Center (NISC)
Clone distribution: NCI-CGAP clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
info@image.llnl.gov
Plate: 21   row: M   column: 21
Seq primer: M13RPl reverse primer (ABI).

FEATURES   Location/Qualifiers
     source   1..608
                  /organism="Xenopus laevis"
                  /db_xref="taxon:8355"
                  /clone_lib="NICHD XGC Ovl"
                  /lab_host="DH10B (phage-resistant)"
                  /note="Organ: ovary; Vector: pCMV-SPORT6; Site_1: NotI;
                  Site_2: SalI; Cloned unidirectionally. Primer: Oligo dT.
                  Average insert size 2.0 kb. Constructed by Life
                  Technologies."

BASE COUNT   162 a   120 c   167 g   159 t
ORIGIN

Qy   344 GACTTGCAGGCCATTGAGCTGGCTGCCCAGAGCAGCAACCATCACTGCCATGCTCAGAAGGGT 403
          || || ||||||||||| |||||||||||||||||||||||||| |||||||| ||||||||
Db   138 GATTTAAACGGTATTGAATTGGGGTACAGAAAACAGCACTGTCATTCTAACCGGGAA 197

Qy   404 CCTGACAGTCACTGTGACCCCAAAGAGAGGGGGAGAAGGGCCAGCTGTGTATGGTAGCCTCT 463
          ||||||||||||||||| |||||| ||||||||| ||||||||||||| |||||||||
Db   198 CAGGGTGATTGGAATAGCAAAGAGAAAAAAAGGCCCGTAGGAGAAGCTGTATGTAGCATCA 257

Qy   464 GCCATCTGCCGTGTTGTTCATGGATCGGAGGAGGAGGTCGTTGGTGGTGCCACACAGCTTG 523
          | ||||||| ||||||||| ||| | |||||||||||| |||||||||||||| |||||
Db   258 GTCGTCTGCCCTACTGTTCATGATCGGAGGAAGTGAATAGGTGGGTACTTGGCACACAGTTTA 317

Qy   524 GCTGTCATGACTGACGGCAGCACACCTGCTCACTGACTTGCCAGCATGCTCATCAGCCTC 583
          || |||||||||||| ||||||||||||||||||||| || ||||||||| ||||||||
Db   318 GCAATCATGACAGATGCTGCTCACCTTCTGACAGACTTTGGTGATGATGATCAGTTTA 377

Qy   584 TTCTCCCTCTGGATGTTCCTCCCGCGCAGCCACCACAGACGAAACTTTGGCAGAGA 643
          || ||||||| |||||||||||||||||||| |||||| ||||||||||||||||
Db   378 TTTGCTCTGTCTGGATGTCCTCAAGGCCCAGCTACTAAAAACTATGAACTTTGGCGTGGCATCGG 437

Qy   644 GCTGAGATCTTGGGAGCCCCTGGTCTCTCGTTACGTGTCCATCTGGGTCGTTGACGGGGTACTG 703
          | ||| |||||||||||||||||||||||| || || || ||||||||| |||||||||||
Db   438 GCAGAGATATTGGGAGGACACTCCCTCTCTGTTGTTCTCTCAATCTGGGTTGTTGACTGGAGTCTTG 497

Qy   704 GTGTACCTTGGCTGTGACGACGCGGCGGACTGATCTCTGGGGACTATGAAAATTGACGGGGACCATG 763
          || |||||||||||||||||||||||||||||||||||||||||||||| ||||||||||||||||
Db   498 GTCTACCTGGCTGTGGGAGCGTATAATATCAGGAGGATTACGAGGATTGCGGGCGCGATGCTATG 557

Qy   764 CTGATCACGTTCGGGCGCTGCCGCCTGTTGGCTGTGTGGCTGTCATTGCCAGTCAATATCAATAATGGGTTAA 812
          |||||||||||||||||||||||||||||||||||||||||||||||||||||| |
Db   558 CTGATCACCTCTCCGGGTGCAGTTGCAGTCAATAATCATAAATGGGTTAA 606

RESULT 15
BI349390
LOCUS   BI349390   619 bp   mRNA   linear   EST 31-JUL-2001
DEFINITION dac60g08.y1 RIKEN Xenopus egg Xenopus laevis cDNA clone
            IMAGE:4434615 5' similar to TR:Q62941 Q62941 ZINC TRANSPORTER
            ZNT-2.;, mRNA sequence.
ACCESSION   BI349390
VERSION   BI349390.1 GI:15043836
KEYWORDS   EST.
SOURCE   African clawed frog.
  ORGANISM Xenopus laevis
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Best Local Similarity   16.5%;   Score 267.4;   DB 10;   Length 524;
                        80.2%;   Pred. No. 1.2e-31;
Matches 344; Conservative   0; Mismatches   71; Indels   14; Gaps   2;

Query Match   15.9%;   Score 257.8;   DB 14;   Length 608;
Best Local Similarity 71.9%;   Pred. No. 3.5e-30;
Matches 337; Conservative   0; Mismatches 132; Indels   0; Gaps   0;

```
REFERENCE     Amphibia; Batrachia; Anura; Mesobatrachia; Pipoidea; Pipidae;
              Xenopodinae; Xenopus.
AUTHORS   1 (bases 1 to 619)
              Clifton,S., Johnson,S.L., Blumberg,B., Song,J., Hillier,L., Pape,D.
              , Martin,J., Wylie,T., Underwood,K., Theising,B., Bowers,Y., Person
              ,B., Gibbons,J., Harvey,N., Ritter,E., Jackson,Y., McCann,R.,
              Waterston,R. and Wilson,R.
TITLE         WashU Xenopus EST project, 1999
JOURNAL       Unpublished (1999)
COMMENT       Contact: Sandy Clifton, Ph.D.
              WashU Xenopus EST project, 1999
              Washington University School of Medicine
              4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
              Tel: 314 286 1800
              Fax: 314 286 1810
              Email: est@watson.wustl.edu
              Library constructed by Piero Carninci (RIKEN, Genome Science
              Laboratory, Japan: http://genome.rtc.riken.go.jp/) using egg RNA
              supplied by Marc Kirschner (Harvard Medical School, Dept. of Cell
              Biology) DNA Sequencing by: Washington University Genome Sequencing
              Center
              Clone distribution: Xenopus clones from this library are available
              through the I.M.A.G.E. Consortium/LLNL at: info@image.llnl.gov
              Trace considered overall poor quality
              High quality sequence stop: 1.
FEATURES             Location/Qualifiers
     source          1..619
                     /organism="Xenopus laevis"
                     /db_xref="taxon:8355"
                     /clone="IMAGE:4434615"
                     /clone_lib="RIKEN Xenopus egg"
                     /tissue_type="egg (unfertilized)"
                     /lab_host="DH10B"
                     /note="Vector: pCS2G (custom); Site_1: BamHI; Site_2: XhoI
                     ; cDNA oligo-dT primed using oligo
                     5'-AGAGAGAGAGAGATCCAATACTGGAGAGTTTTTTTTTTTTTTTTTVN-3' and
                     directionally cloned into XhoI and BamHI sites using the
                     5' adaptor
                     5'-AGAGAGAGAGCTCGAGCTCTAATAAGGTGGACTATAGAACCA-3'. Size
                     selection was done automatically by the cloning vector
                     (lambda FLC-II, P. Carninci et al., unpublished). Library
                     was amplified at the phage stage and normalized (RoT
                     value = 5). This library was constructed using the
                     CapTrapper method and Trehalose thermoactivated reverse
                     transcriptase to enrich for full-length clones. Library
                     constructed by Piero Carninci (RIKEN, Genome Science
                     Laboratory, Japan: http://genome.rtc.riken.go.jp/) using
                     egg RNA supplied by Marc Kirschner (Harvard Medical
                     School, Dept. of Cell Biology). References: Methods
                     Enzymol.303:19-44 and Genome Res. 10:1617-1630 Please
                     contact Marc Kirschner (marc@hms.harvard.edu) for
                     information on obtaining aliquots of this cDNA library."
BASE COUNT      165 a      141 c      137 g      176 t
ORIGIN
```

```
Query Match   15.1%;  Score 244.8;  DB 13;  Length 619;
Best Local Similarity  64.2%;  Pred. No. 3.3e-28;
Matches 397;  Conservative  0;  Mismatches 197;  Indels  24;  Gaps  1;

Qy     695 GGGGTACTGGTGTACCTGGGCTGTGGACGGGCGTGATCTCTGGGGACTATGAAATTGACCGG 754
           || || ||||  || | ||||||||||||| || | ||||||  ||||| ||| | ||||||
Db       2 GGAGTCTGGCCTACCTGGCCTGAGGACCGTATAAATATCAGAGAATTACTATATTGCGGGC 61

Qy     755 GGGACCAATGCTGATCACGTCGGGCGTCGCGCTTGGGCGTGTGACATCATTAATGGGGTTGACC 814
           ||||| |||||||||||  || | | |||||| |||||| || | |||| |||| |||||||||
Db      62 GATGCTATGCTGATCACCTCTCCGGGGCAGTTGCAGCCAATATCAATAATGGGTTAATT 121

Qy     815 CTTCACCAGTCTGGCCATGGGGCCATGCCACACCGCCACCAACCAGC------------- 860
           |||||||| ||| |||| ||| || | |||| |||| |||| ||||||
Db     122 CTTCATCATACAAGGGCATGGCACACACTCCCATTCCATGTGATCGGCGCAT 181

Qy     861 -----------AGGAGGAGAACCCCAGCGTCCGAGCTCGCTTCATCCACATGTGATCGGCCAC 910
           |   | |||||| ||||||||||||||| |||||||||||||| ||||
```

```
Db     602 CACACCACTACTATCCAG 619
           |||| |||| | ||||||
Qy    1271 CACACCGTGACCATCCAG 1288

Db     542 GCAGACTCGCAAAACAGTGCTAAAAAACGTCAGCTCACACTACAGAATAAGTTCCACTTT 601
           |||||||||  |||  | || ||||   || ||||| |||||  ||||||| |||||
Qy    1211 ACAGAGCGCCCAGGCTGTGCTGAAGACAGCCAGCACGCCGCCTCAAGGAGGAAGTTCCACTTC 1270

Db     482 ATCTGGGACCTGACAGTATCCAACGTGTCCTCTTCATTATAGCAATCAATCAGAAT 541
           | | |||||| |||| ||  |||| | | | ||||    ||  ||||    || ||
Qy    1151 ATCTGGCGCACTGAGGGTGGGCCCAGCCTGTTTCGTCGTCTCCACATCGCCATTGCTCAGAAT 1210

Db     422 AACTTGGTGAAGGACACCTTGCTCTCGATCAGAGGAGTGAAAAGCTTCTTCATTAGTCTGCAT 481
           ||||||||| ||  | |  |||| || || |||| ||| |||| ||| || |||| || ||||
Qy    1091 ACAGCTGTTCGTGATCTGCTGCTGTCGGTGGAGGGGTAGAAGCCCTGCACAGCCTGCAT 1150

Db     362 ACCATCCTTAGAGATGTCCTTCTAGTCCTTAATGGAAGGAAGACCCTTAAGGGAGTGGGATTTT 421
           ||||||| ||||| | ||  || || ||  | ||| |||| ||||| ||| | |||| ||||
Qy    1031 ACCATCCTGAGAGATGTGATCCTGGTGTTGGATGGAAGGGGACCCCCAAGGGCGTTGGACTTC 1090

Db     302 AAAATTATTGATCCTATAATCCACATTCCTGTATTCTGTGCTGAGTTCTGATCACACACTG 361
           | ||  | |||   | |||  || || ||  |||| |  |  ||||| | |  || |
Qy     971 AAGTATGTAGACCCCATCTGCACCTTCGTCTTCTCCCATCCTGGTCCTGGGGACAACCTTG 1030

Db     242 TTATTACTAAGCCGTTGGGGTCATTGGTCCTAGCCCTACGTAATTTATTATAGGCCGGAATAT 301
           ||| | || ||| | ||  || | | |||||||  | | |  ||| || || ||| ||| |
Qy     911 TTTATGCAGAGACGATGGGGTGTCCTAGTGCAGGCCAGCCTATATTTTTATACTTCAAGCCATGGCTT 970

Db     182 GGTGCTGAGATCATGGTAATCCTTATTAGTGTCCGTGCTGCTGCTTTTATCCATGTGGACGGGGGAT 241
```

THIS PAGE BLANK (USPTO)

GenCore version 5.1.4_p5_4578
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 29, 2003, 16:26:54 ; Search time 75.0846 Seconds
                         (without alignments)
                         6604.503 Million cell updates/sec

Title: US-09-691-219-1
Perfect score: 1617
Sequence: 1 cgcggtgaacagcacgaggg.........aaaaaaaaaaaaaaaaatgtt 1617

Scoring table: IDENTITY_NUC
              Gapop 10.0 , Gapext 1.0

Searched: 441362 seqs, 153338381 residues

Total number of hits satisfying chosen parameters: 882724

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
                 Maximum Match 100%
                 Listing first 45 summaries

Database :   Issued_Patents_NA:*
              1: /cgn2_6/ptodata/1/1na/5A_COMB.seq:*
              2: /cgn2_6/ptodata/1/1na/5B_COMB.seq:*
              3: /cgn2_6/ptodata/1/1na/6A_COMB.seq:*
              4: /cgn2_6/ptodata/1/1na/6B_COMB.seq:*
              5: /cgn2_6/ptodata/1/1na/PCTUS_COMB.seq:*
              6: /cgn2_6/ptodata/1/1na/backfiles1.seq:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result NO.	Score	% Query Match	Length	DB	ID	Description
1	143.8	8.9	1551	4	US-09-461-474-7	Sequence 7, Appli
2	132	8.2	1873	4	US-09-461-474-9	Sequence 9, Appli
3	118.2	7.3	790	4	US-09-461-474-13	Sequence 13, Appli
4	106.2	6.6	1208	4	US-09-461-474-11	Sequence 11, Appli
5	68.8	4.3	504	4	US-09-134-001C-30	Sequence 30, Appli
6	61.8	3.8	903	5	PCT-US95-06406A-21	Sequence 21, Appl
7	58.6	3.6	2378	4	US-08-802-805D-20	Sequence 20, Appl
8	57	3.5	2269	4	US-09-394-645-1	Sequence 1, Appl
9	57	3.5	2269	4	US-09-243-560B-1	Sequence 1, Appli
10	56.4	3.5	2719	3	US-08-706-216-1	Sequence 1, Appli
11	55.8	3.5	789	4	US-09-020-956-32	Sequence 32, Appl
12	55.8	3.5	789	4	US-09-030-607-32	Sequence 32, Appl
13	55.8	3.5	789	4	US-09-605-785-32	Sequence 32, Appl
14	55.8	3.5	789	4	US-09-439-313-32	Sequence 32, Appl
15	55.8	3.5	789	4	US-09-352-616A-32	Sequence 32, Appl
16	55.8	3.5	789	4	US-09-232-149A-32	Sequence 32, Appl
17	55.8	3.5	8438	1	US-07-945-283-1	Sequence 1, Appli
18	55.2	3.5	1069	4	US-09-372-42A-7	Sequence 7, Appli
19	54.8	3.4	1153	4	US-09-149-476-41	Sequence 41, Appl
20	54.4	3.4	1359	4	US-09-387-574-11	Sequence 11, Appl
21	54.4	3.4	1359	4	US-09-668-09G-11	Sequence 11, Appl
22	54.2	3.4	1509	4	US-09-149-476-179	Sequence 179, App
23	54.2	3.4	2665	4	US-08-97-089-5	Sequence 5, Appli
24	54	3.4	976	2	US-08-504-459-9	Sequence 9, Appli
25	54	3.3	1206	4	US-09-465-558-53	Sequence 53, Appl
26	54	3.3	1578	4	US-09-416-050A-1	Sequence 1, Appli
27	54	3.3	1578	4	US-09-664-800-1	Sequence 1, Appl
28	54	3.3	1578	4	US-09-665-309-1	Sequence 1, Appli
29	54	3.3	1578	4	US-09-661-569-1	Sequence 1, Appli
30	53.6	3.3	1193	4	US-09-372-422A-23	Sequence 23, Appl
31	53.6	3.3	1604	1	US-08-665-966-9	Sequence 9, Appli
32	53.6	3.3	1604	1	US-09-041-780-9	Sequence 9, Appli
33	53.6	3.3	2246	4	US-09-363-708-3	Sequence 3, Appli
34	53.4	3.3	1013	1	US-07-920-519-30	Sequence 30, Appli
35	53.4	3.3	1013	1	US-08-086-410-23	Sequence 23, Appl
36	53.4	3.3	1013	1	US-08-314-566-30	Sequence 30, Appl
37	53.4	3.3	9704	4	US-09-814-951A-3	Sequence 3, Appli
38	53.2	3.3	1066	1	US-08-157-101A-4	Sequence 4, Appli
39	53.2	3.3	1910	4	US-08-974-691-7	Sequence 7, Appli
40	53.2	3.3	1929	4	US-09-146-950-1	Sequence 1, Appli
41	53	3.3	857	1	US-08-308-883-1	Sequence 1, Appli
42	53	3.3	857	1	US-08-730-163-1	Sequence 1, Appli
43	53	3.3	857	4	US-08-256-799-1	Sequence 1, Appli
44	53	3.3	857	4	US-08-462-437-1	Sequence 1, Appli
45	53	3.3	3080	4	US-09-099-041A-25	Sequence 25, Appl

ALIGNMENTS

RESULT 1
US-09-461-474-7
; Sequence 7, Application US/09461474
; Patent No. 6278042
; GENERAL INFORMATION:
; APPLICANT: Allen, Steve
; APPLICANT: Rafalski, Antoni
; APPLICANT: Sakai, Hajime
; TITLE OF INVENTION: Plant Metal Transporters
; FILE REFERENCE: BB1303 US NA
; CURRENT APPLICATION NUMBER: US/09/461,474
; CURRENT FILING DATE: 1999-12-14
; EARLIER APPLICATION NUMBER: 60/112,562
; EARLIER FILING DATE: 1998-12-16
; NUMBER OF SEQ ID NOS: 17
; SOFTWARE: Microsoft Office 97
; SEQ ID NO 7
; LENGTH: 1551
; TYPE: DNA
; ORGANISM: Zea mays
US-09-461-474-7

Query Match   8.9%; Score 143.8; DB 4; Length 1551;
Best Local Similarity 57.4%; Pred. No. 5.2e-22;
Matches 259; Conservative 0; Mismatches 192; Indels 0; Gaps 0;

```
Qy    854 AACCAGCAGGAGGAGGAGAACCCCAGCGTCCGAGCTGCCTTCATCCATGTGATCGGGGACTTT 913
              | |||| |||      ||| ||||| || |||| ||||| ||||||||  ||||
Db    899 AAGAAGCCTCGGCGGGGAACATCACAGTGCACAGCCGCCATACCTCCACGTGCTCGGGGACTCC 958

Qy    914 ATGCAGAGCATGGGTGTCCTAGTGGCAGCCTATATTTATACTTCAAGCCAGAATACAAG 973
              |  ||||||||||||||| ||  |  ||  ||     |   ||||||  ||||| |||
Db    959 GTCCAGAGGCGTCGGGGGTCATGGGTGGCGGGGGATCATCGGTACAAGCCGGAGTGGAAG 1018

Qy    974 TAIGTAGAGCCCCATCTGCGACCTTCGTTCTTCTCCATCCTTCGGGGGACAACCTTGACC 1033
              | ||||||||||||||| |||||| | ||| |||||| | ||||||||||||||||
Db   1019 GTCATCGACCTCATCTGCACGCTCGTCTTCTCCGTGGTGGTGCGTTCACCACGATCCGG 1078

Qy   1034 ATCCTTGAGAGATGTTGATCCTTGGTGTTGATGGAAGGGACCCCCAAGGGGCGTTGACTTCACA 1093
              ||  | ||||| ||||||||||||||||||||||||| ||||||||  ||| || ||| |||
Db   1079 ATGCTTGCCCAGCATCCTCGAAAGTGTTGATGGAGAGCACGCCCCGCGAGATCGAGCGCACC 1138

Qy   1094 GCTGTTTCGTGCATCTGTCTTGCTTCGGTGGAGGGGGTAGAAGCCCTGCACAGCCTGCATATC 1153
              |||       |||      | |||||||||| ||| |||||||    | |||| ||||||||
Db   1139 AGGCTGGAGAGCGGGCGCTCTGCCGGGGATGGAGGGCCGTGGTGGCCGTCCACGAGCTCCACATC 1198

Qy   1154 TGGGCACTGACGGTGGGCCCAGCCGCCTGTTCGTTCTTCTGTCTCCACATCGCCATTGCTCAGAATACA 1213
              ||||||| ||||||||||| ||||||||||| | || | |||| |||||| ||||| || ||||||
Db   1199 TGGGCCATCACGGTGGGCAAGGTGCTGCTGCGTGCCCANGTGACCAATCGCGCCAGGAATGCG 1258
```

Left column

```
Qy   1214  GACGCCCAGCGTGTGCTGAAGACACAGCCAGACGCCGCCTCCAAAGGGAAGTTCCACTTCCAC  1273
           ||||| || | || || | || | ||||||| ||||| |||||| ||||| |||| || 
Db   1259  GACGCTGATGAGATCCTTGACAAGGTGATCGGGTACATCAAGAGACGGAGTACAACAATCAGC  1318

Qy   1274  ACCGTGACCATCCAGATCGAGGACTACTCGG  1304
           |||||||||||||| ||||| |||| || || 
Db   1319  CATGTGACCATCCAGGTCGAGCGCGCGAGTAGG  1349
```

RESULT 2
US-09-461-474-9
Sequence 9, Application US/09461474
Patent No. 6278042
; GENERAL INFORMATION:
; APPLICANT: Allen, Steve
; APPLICANT: Rafalski, Antoni
; APPLICANT: Sakai, Hajime
; TITLE OF INVENTION: Plant Metal Transporters
; FILE REFERENCE: BB1303 US NA
; CURRENT APPLICATION NUMBER: US/09/461,474
; CURRENT FILING DATE: 1999-12-14
; EARLIER APPLICATION NUMBER: 60/112,562
; EARLIER FILING DATE: 1998-12-16
; NUMBER OF SEQ ID NOS: 17
; SOFTWARE: Microsoft Office 97
; SEQ ID NO 9
; LENGTH: 1873
; TYPE: DNA
; ORGANISM: Oryza sativa
US-09-461-474-9

Query Match   8.2%;   Score 132; DB 4; Length 1873;
Best Local Similarity   56.4%;   Pred. No. 1.8e-19;
Matches 246; Conservative 0; Mismatches 190; Indels 0; Gaps 0;

```
Qy   869  AACCCCAGGCGTCCGAGCTGCCTTCATCCATCGTGATCGGGCGACTTTATGCAGAGCATGGGT  928
Db   992  AATATCAATGTACACAGTGCTTATCYGTCACATGCTTGGGATTCAAATCAACCAGAGCATCGGT  1051

Qy   929  GTCCTAGTGGGCAGCCTATATATTTTATACTTCAAGCCAGAATACAGTATGTAGAGACCCCATC  988
Db  1052  GTGATGATTGGGAGGGGCTATCATCTGGTACAAGCCGAGTGGAAGATTATTGATCTCATC  1111

Qy   989  TGCACCTTCGTGTCTTCTCCATCCTGGTCTGGGGACAAACCTTGACCATCCTGAGAGATGTG  1048
Db  1112  TGCACCCTCATCTTCTCCGTGATCGTGACTCCTTCACCAACATCAAAGATGCTCGCCAACATC  1171

Qy  1049  ATCCTGGTGTTGATGGGAAGGGACACCCCCAAGGGCGTTGACTTCACAGCTGTTCGTGATCTG  1108
Db  1172  CTTGAGGTCCTGATGGAGAGCACGCCCCGCGAGATGCAGAGTACCAGCCTTGAGAATGGC  1231

Qy  1109  CTGCTCTCGGGTGCGAGGGGGTACGAAGCCCTGCACAGCGCCTGCATAATCTGGGCACTGACGGTG  1168
Db  1232  CTCCGCGACATGGAGCGTGGTTGCAGTACAATGAGCCTGCACATCTTGGCCCCATAACCGGTG  1291

Qy  1169  GCCCACCTGTTCTGTGTCTGCACATCGCCATTGCTCAGAATACAGACGCCCAGGCTGTG  1228
Db  1292  GGGAAGTTCTCCTGGCGTGCCATGTGACAAATCACTCAGGACACGCCTGATCAAATG  1351

Qy  1229  CTGAAGCAGCCAGCAGCGCCTCCAAGGGAAGTTCCACTTCCACACCGTGACCATCCAG  1288
Db  1352  CTGGACAAGGTGATTGGGTACATCAAGCTCTGAGTAGTACAACATCAGCCATTGACCATTCAG  1411

Qy  1289  ATCGAGGACTACTCGG  1304
Db  1412  ATTGAGCGCGAGTAGG  1427
```

Right column

; APPLICANT: Allen, Steve
; APPLICANT: Rafalski, Antoni
; APPLICANT: Sakai, Hajime
; TITLE OF INVENTION: Plant Metal Transporters
; FILE REFERENCE: BB1303 US NA
; CURRENT APPLICATION NUMBER: US/09/461,474
; CURRENT FILING DATE: 1999-12-14
; EARLIER APPLICATION NUMBER: 60/112,562
; EARLIER FILING DATE: 1998-12-16
; NUMBER OF SEQ ID NOS: 17
; SOFTWARE: Microsoft Office 97
; SEQ ID NO 13
; LENGTH: 790
; TYPE: DNA
; ORGANISM: Triticum aestivum
US-09-461-474-13

Query Match   7.3%;   Score 118.2; DB 4; Length 790;
Best Local Similarity   54.2%;   Pred. No. 1.3e-16;
Matches 240; Conservative 0; Mismatches 203; Indels 0; Gaps 0;

```
Qy   852  CCAACCACCAGGAGGAGAACCCCAGGCGTCCGAGCTGCCTTCATCCATCGTGATCGGGACT  911
Db   911  ...

Qy  1032  CCATCCTGAGAGATGTGATCCTGGTGTTGATGGGAAGGGACACCCCCAAAGGGCGGTTGACTTCA  1091
Db  1091  ...

Qy  1152  TCTGGGCACTGACGGTGCGCCAGCCGTTCTGTGTCGTGCTCACACATCGCCATTGCTCAGAATA  1211
Db  1211  ...

Qy  1272  ACACCGTGACCATCCAGATCGAG  1294
Db   445  GTCATGTGACCATTCAGATTGAG  467
```

RESULT 4
US-09-461-474-11
Sequence 11, Application US/09461474
Patent No. 6278042
; GENERAL INFORMATION:
; APPLICANT: Allen, Steve
; APPLICANT: Rafalski, Antoni
; APPLICANT: Sakai, Hajime
; TITLE OF INVENTION: Plant Metal Transporters
; FILE REFERENCE: BB1303 US NA
; CURRENT APPLICATION NUMBER: US/09/461,474
; CURRENT FILING DATE: 1999-12-14
; EARLIER APPLICATION NUMBER: 60/112,562
; EARLIER FILING DATE: 1998-12-16
; NUMBER OF SEQ ID NOS: 17
; SOFTWARE: Microsoft Office 97
; SEQ ID NO 11
; LENGTH: 1208
; TYPE: DNA

```
;      ORGANISM: Glycine max
US-09-461-474-11

Query Match          6.6%;  Score 106.2;  DB 4;  Length 1208;
Best Local Similarity  52.1%;  Pred. No. 5.2e-14;
Matches  237;  Conservative   0;  Mismatches  218;  Indels   0;  Gaps   0;

Qy   845 GGCACCACCAACCACGAGGAGGAGGAGAACCCAGCGTCCGAGCTGCCTTCATCCATGTGATC 904
             |||||   |||||||  ||||   ||| ||  ||||||||||||  ||| |||
Db   595 GGAACTAAGAGAAGAAGCAATGGGACACATAAAATGACAGAGGGGTCTTATCTCCATGTTCTT 654

Qy   905 GGCGACTTATGCAGAGAGCATGGGTGGTGTTCCTAGTGGCAGCCTATATTTTATACTTCAAGCCA 964
             ||| |||  |||   |  ||| |||  || ||| ||| ||||   |||
Db   655 GGGGACTTCATCTATCCAAAGTATTGGGGTAATGAATTCGGGGAGCAGTCATATGGTATAACCT 714

Qy   965 GAATACAAAGTATGTAGAGACCCCATCTACCACCTTCGTCTTCTTCATCCTCTGGGGACA 1024
             ||| || |||||  ||||||||  |||  || ||   ||||||||  ||||| |||
Db   715 CGTTGGCAAATTGTTGATTTAATTCTCGCACTCTAATCTTTTCAGTAATTGTTATTGGGGACA 774

Qy  1025 ACCTTGACCATCCTGAGAGATGTGATCCTCGTTGTTCATGAAGGGGACCCCAAGGGGGTT 1084
             |||  || ||| |||   ||||||| ||   ||   ||||||| || ||  |||||
Db   775 ACCATCAACAATGCTGGGAAACATTTTGGAGTCCTCGTTGGTGAGAACACACCTCGTGAGATA 834

Qy  1085 GACTTCACAGCTGTTCGTGATCGTGCTCGTCGTCGGGGTGGAGGGGGTAGAAGCCCTGCACAGC 1144
             || |||||||| | |||||||||  | ||  ||| ||| ||| ||||||  |||| || ||
Db   835 GATGCTACTAAGCTTGAAAGGGGGCGTGTTGGATATGGAAGATGGTAGTAGTGGGCCCTTCATGAA 894

Qy  1145 CTGCATATCTGGGCACTGACGGTGGGCCCACCGCTGTCTGTCTGTGTCACATCGCCATTGCT 1204
             |||||| |  |||||| ||    ||| ||||| |  ||  || ||  |||||| | ||
Db   895 CTGCATATATGGGCCATTACACTGGGAAAGGTTTTTCGCTTGCATGTCATGTTAAGATCAGA 954

Qy  1205 CAGAATACACAGACGCCCAGGCTGTGCTGAAGACAGCAGCCAGCAGCGCGCCTCCAAGGGAAGTTC 1264
             |||||| | ||  ||||| |||||||||  ||||| || ||| | | ||||||  |||||| |||
Db   955 CGTGAAGCAGGATGCAGACTTGGTGCTGCGACAAAGGTTATAGACTATATCAAAAGGGTTTAT 1014

Qy  1265 CACTTCCACACCGTGACCATCCAGATCGATCGTGAGGACTA 1299
             || |  || |  |||||||||||| |||||| ||||| ||||
Db  1015 AACATCAGCCATGTCACCATACAGATAGAGCGTTA 1049
```

```
RESULT 5
US-09-134-001C-30
Sequence 30, Application US/09134001C
;      Patent No. 6380370
;      GENERAL INFORMATION:
;      APPLICANT: Lynn Doucette-Stamm et al
;      TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCCUS
;      TITLE OF INVENTION: EPIDERMIDIS FOR DIAGNOSTICS AND THERAPEUTICS
;      FILE REFERENCE: GTC-007
;      CURRENT APPLICATION NUMBER: US/09/134,001C
;      CURRENT FILING DATE: 1998-08-13
;      PRIOR APPLICATION NUMBER: US 60/064,964
;      PRIOR FILING DATE: 1997-11-08
;      PRIOR APPLICATION NUMBER: US 60/055,779
;      PRIOR FILING DATE: 1997-08-14
;      NUMBER OF SEQ ID NOS: 5674
;      SEQ ID NO 30
;      LENGTH: 504
;      TYPE: DNA
;      ORGANISM: Staphylococcus epidermidis
;      FEATURE:
;      NAME/KEY: unsure
;      LOCATION: (471)
;      OTHER INFORMATION: Identity of nucleotide at the above locations are unknown.
US-09-134-001C-30

Query Match          4.3%;  Score 68.8;  DB 4;  Length 504;
Best Local Similarity  52.4%;  Pred. No. 3.9e-06;
Matches  176;  Conservative   0;  Mismatches 157;  Indels   3;  Gaps   1;

Qy   871 CCCCAGGCGTCCGAGCTGCCTTCATCCATGTGATCGGCGACTTTATGCAGAGCATGGGTGCT 930
             |||   |  || | ||| ||||| |||||||  ||| ||  |||||| |||| |||||
Db     6 CCTCAGCACGCGGCGCGCTGCCCTCGTCGTCATGTGATGGGAGACTTGCTCGGTCGGCGG 65
```

```
RESULT 6
PCT-US95-06406A-21
Sequence 21, Application PC/TU950 6406A
;      GENERAL INFORMATION:
;      APPLICANT: Janet D. Robishaw, Charles Kunsch
;      TITLE OF INVENTION: cDNA Clones Encoding Human G Protein
;      TITLE OF INVENTION: Subunits
;      NUMBER OF SEQUENCES: 23
;      CORRESPONDENCE ADDRESS:
;      ADDRESSEE:
;      STREET:
;      CITY:
;      STATE:
;      COUNTRY:
;      ZIP:
;      COMPUTER READABLE FORM:
;      MEDIUM TYPE: DISKETTE, 3.5 INCH, 1.44 Mb STORAGE
;      COMPUTER: IBM 486
;      OPERATING SYSTEM: WINDOWS FOR WORKGROUPS
;      SOFTWARE: WORDPERFECT 5.1
;      CURRENT APPLICATION DATA:
;      APPLICATION NUMBER: PCT/US95/06406A
;      FILING DATE: Herewith
;      CLASSIFICATION:
;      PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:
;      FILING DATE:
;      ATTORNEY/AGENT INFORMATION:
;      NAME:
;      REGISTRATION NUMBER:
;      REFERENCE/DOCKET NUMBER:
;      TELECOMMUNICATION INFORMATION:
;      TELEPHONE:
;      TELEFAX:
;      INFORMATION FOR SEQ ID NO: 21:
;      SEQUENCE CHARACTERISTICS:
;      LENGTH: 903
;      TYPE: Nucleic Acid
;      STRANDEDNESS: Single
;      TOPOLOGY: Linear
;      ANTI-SENSE: NO
PCT-US95-06406A-21

Query Match          3.8%;  Score 61.8;  DB 5;  Length 903;
Best Local Similarity  71.7%;  Pred. No. 0.00014;
Matches  81;  Conservative   0;  Mismatches  32;  Indels   0;  Gaps   0;

Qy  1501 TCCCAGGCCTCCCATCTGACTACAGCCACAGGGTGGGGGACTCAGCGGGGTAATAAGGCTTAGTGTG 1560
             ||||||| |  ||||||| || | ||  ||||| | |||||||   || || ||| ||||||||
Db   746 TCTCAGTCTCACCTGCGAGCTACTGGGAGGGGTAAAGCCCATTTGAAGAATAAAAGTCATCCAG 805
```

```
Qy   931 CCTAGTGGCGCAGCCTATATTTTATACTTCAAGCCAGAAATACAAGTATGTAGAGACCCCATCTG 990
             ||||| |||  ||||||||  |||| ||||||||| | ||||  | |||| |||| || |||
Db    66 CATCATCCGCGCCGGCCGTGATTTACTGTA---CAGGCTGGATGCAGATCGAGATCGACCCGCTGCT 122

Qy   991 CACCTTCGTCTTCTTCCATCCTGGTCCTGGGGACAACAACCTTGACCATCCTGAGAGATGTGAT 1050
             |||||  || |  ||  ||||   || |||  |||||||| || |||  ||||  |||  |
Db   123 GTCCGGTGCGTGGGGTCCCTCGCTCCTGATTTAAGTCGACGTTTGGGCGTACTCGCGGAGTCCTTA 182

Qy  1051 CCTGGTTGTTGATGGAAGAGGAGACCCCCAAGGGGGTGTTGACTTCACACGTGTTCGTGATCTGCT 1110
             ||| | ||   ||||| | |||||||| |||||| || |  | ||| | ||| |  |||||
Db   183 TCACTTCCCTGATGGAAGCGGCGCCTGCCCCCCTTGACATCGACTACATCGTGGCCACCGTGGCCT 242

Qy  1111 GCTGTCGGTGGGAGGGGGGTAGAAGCCCTGCACAGCCTGCATATCTGGGGCACTGCACGTGTTGCC 1170
             ||| ||| | ||  | ||||  ||||||  ||||| || | |||  ||| ||| | |||  |||
Db   243 GGATCAGGTGGGACGGGCGTGATTGCCGTGATTGCCGTGGCACTGACGGTGGC 302

Qy  1171 CCAGGCTGTTTCTGTGTCGTCCACATCGCCATTGCTCA 1206
             ||||||||| |  |||  | |||| |||  |||  |||
Db   303 CCAACCAGCCCTGATCGGCGGCCACGTGGGAAATCGAACA 338
```

Qy   1561 ACCCTGAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 1613
          |||| ||||||||||||||||||||||||| |||||||||||||
Db   806 AGCCTCAAAAAAAAAAAAAAAAAAAAAAAAAAAATAAAAAAAAAAAAAAAAAAAAAAAAA   858

RESULT 7
US-08-802-805D-20
Sequence 20, Application US/08802805D
Patent No. 6372453
GENERAL INFORMATION:
  APPLICANT: Robert D. Klein
  TITLE OF INVENTION: Neurturin Receptor
  NUMBER OF SEQUENCES: 28
  CORRESPONDENCE ADDRESS:
    ADDRESSEE: Genentech, Inc.
    STREET: 1 DNA Way
    CITY: South San Francisco
    STATE: California
    COUNTRY: USA
    ZIP: 94080
  COMPUTER READABLE FORM:
    MEDIUM TYPE: 3.5 inch, 1.44 Mb floppy disk
    COMPUTER: IBM PC compatible
    OPERATING SYSTEM: PC-DOS/MS-DOS
    SOFTWARE: WinPatin (Genentech)
  CURRENT APPLICATION DATA:
    APPLICATION NUMBER: US/08/802,805D
    FILING DATE: 18-Feb-1997
    CLASSIFICATION: 536
  ATTORNEY/AGENT INFORMATION:
    NAME: Torchia, PhD., Timothy E.
    REGISTRATION NUMBER: 36,700
    REFERENCE/DOCKET NUMBER: P1086
  TELECOMMUNICATION INFORMATION:
    TELEPHONE: 650/225-8674
    TELEFAX: 650/952-9881
INFORMATION FOR SEQ ID NO: 20:
  SEQUENCE CHARACTERISTICS:
    LENGTH: 2378 base pairs
    TYPE: Nucleic Acid
    STRANDEDNESS: Single
    TOPOLOGY: Linear
US-08-802-805D-20

Query Match   3.6%; Score 58.6; DB 4; Length 2378;
Best Local Similarity 62.8%; Pred. No. 0.0009;
Matches 91; Conservative 0; Mismatches 54; Indels 0; Gaps 0;

Qy   1469 TGACCTCTGCCCCACTCCAGGAATGGAGCTCTTCCCAGCCTCCCATCTGACTACAGCCAG 1528
          || || ||| ||||| |||||||| ||| | |||| ||||| | | | || |
Db   2156 TGGCCCTGCCACAGCTCCCGGATTGTTGATGGGCACTCCCTGACCTTTGCTTGTGAGTGAG 2215

Qy   1529 GGTGGGGACTCAGCGGGTATAAAGCTAGTGTGACCCTGAAAAAAAAAAAAAAAAAAAAAAA 1588
          ||| || ||| ||| |
Db   2216 AAGCTGAATGTAGCTGAAAATCAACTCTTCTTACACTTAAAAAAAAAAAAAAAAAAAAAAA 2275

Qy   1589 AAAAAAAAAAAAAAAAAAAA 1613
          ||||||||||||||
Db   2276 AAAAAAAAAAAAA 2300

RESULT 8
US-09-394-645-1
Sequence 1, Application US/09394645
Patent No. 638371
GENERAL INFORMATION:
  APPLICANT: Sassetti, Christopher M.
  APPLICANT: Rosen, Steven R.
  TITLE OF INVENTION: Endoglycan: A NOVEL PROTEIN HAVING SELECTIN
  TITLE OF INVENTION: LIGAND AND CHEMOKINE PRESENTATION ACTIVITY
  FILE REFERENCE: 6510-122US1
  CURRENT APPLICATION NUMBER: US/09/394,645
  CURRENT FILING DATE: 1999-09-13
  NUMBER OF SEQ ID NOS: 2
  SOFTWARE: FastSEQ for Windows Version 3.0
SEQ ID NO 1
  LENGTH: 2269
  TYPE: DNA
  ORGANISM: human
US-09-394-645-1

Query Match   3.5%; Score 57; DB 4; Length 2269;
Best Local Similarity 60.8%; Pred. No. 0.002;
Matches 93; Conservative 0; Mismatches 60; Indels 0; Gaps 0;

Qy   1461 CCCCTCCTGACCTCTGCCCCACTCCAGGAATGGAGCTCTTCCCAGCCTCCCATCTGACT 1520
          || | || |||||| ||| | ||||| || |||||| |||||||||| ||| |
Db   2064 CCGCCCCTGAACCCCGGCCCCGCGGGGCGGCCGCTTCCTGCGCCCCGGGGGGACTCAAT 2123

Qy   1521 ACAGCCAGGGTGGGGACTCAGCGGGGTATAAAGCTAGTGTGACCCTGAAAAAAAAAAAA 1580
          | | ||||| ||| | ||| | ||| || ||||||| |||||||||||||||
Db   2124 TAAACCCGGCCCCGGGGAGACCACGCGGGGCCCAGCGAAAAAAAAAAAAAAAAAAAAAA 2183

Qy   1581 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 1613
          |||||||||||||||||||||||||
Db   2184 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 2216

RESULT 9
US-09-243-560B-1
Sequence 1, Application US/09243560B
Patent No. 6395882
GENERAL INFORMATION:
  APPLICANT: Rosen, Steven D.
  APPLICANT: Sassetti, Christopher M.
  TITLE OF INVENTION: No. 6395882e1 Selectin Ligands
  FILE REFERENCE: UCAL097US1
  CURRENT APPLICATION NUMBER: US/09/243,560B
  CURRENT FILING DATE: 1999-02-03
  PRIOR APPLICATION NUMBER: US 60/074,389
  PRIOR FILING DATE: 1998-02-11
  NUMBER OF SEQ ID NOS: 10
  SOFTWARE: FastSEQ for Windows Version 4.0
SEQ ID NO 1
  LENGTH: 2269
  TYPE: DNA
  ORGANISM: Homo sapiens
US-09-243-560B-1

Query Match   3.5%; Score 57; DB 4; Length 2269;
Best Local Similarity 60.8%; Pred. No. 0.002;
Matches 93; Conservative 0; Mismatches 60; Indels 0; Gaps 0;

Qy   1461 CCCCTCCTGACCTCTGCCCCACTCCAGGAATGGAGCTCTTCCCAGCCTCCCATCTGACT 1520
          || | || |||||| ||| | ||||| || |||||| |||||||||| ||| |
Db   2064 CCGCCCCTGAACCCCGGCCCCGCGGGGCGGCCGCTTCCTGCGCCCCGGGGGGACTCAAT 2123

Qy   1521 ACAGCCAGGGTGGGGACTCAGCGGGGTATAAAGCTAGTGTGACCCTGAAAAAAAAAAAA 1580
          | | ||||| ||| | ||| | ||| || ||||||| |||||||||||||||
Db   2124 TAAACCCGGCCCCGGGGAGACCACGCGGGGCCCAGCGAAAAAAAAAAAAAAAAAAAAAA 2183

Qy   1581 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 1613
          |||||||||||||||||||||||||
Db   2184 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 2216

RESULT 10
US-08-706-216-1
Sequence 1, Application US/08706216
Patent No. 6140098
GENERAL INFORMATION:
  APPLICANT: Balasubramanian, Sriram
  APPLICANT: Ford, John
  APPLICANT: Gorman, Daniel M.
  APPLICANT: Zurawski, Gerard
  TITLE OF INVENTION: MAMMALIAN PROTEASES; RELATED REAGENTS

NUMBER OF SEQUENCES: 6
CORRESPONDENCE ADDRESS:
    ADDRESSEE: DNAX Research Institute
    STREET: 901 California Avenue
    CITY: Palo Alto
    STATE: California
    COUNTRY: USA
    ZIP: 94304-1104
COMPUTER READABLE FORM:
    MEDIUM TYPE: Floppy disk
    COMPUTER: IBM PC compatible
    OPERATING SYSTEM: PC-DOS/MS-DOS
    SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
    APPLICATION NUMBER: US/08/706,216
    FILING DATE: 30-AUG-1996
    CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
    NAME: Ching, Edwin P.
    REGISTRATION NUMBER: 34,090
    REFERENCE/DOCKET NUMBER: DX0613
TELECOMMUNICATION INFORMATION:
    TELEPHONE: 415-852-9196
    TELEFAX: 415-496-1200
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
    LENGTH: 2719 base pairs
    TYPE: nucleic acid
    STRANDEDNESS: single
    TOPOLOGY: linear
MOLECULE TYPE: cDNA
FEATURE:
    NAME/KEY: CDS
    LOCATION: 337..2541
US-08-706-216-1

Qy   1548 TAAAGCTAGTGTGACCCTGAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 1607
           |||||||| ||| |||AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA
Db   2638 TAAAGCTACCGGGCCACCTTAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 2697

Qy   1608 AAAAAA 1613
           ||||||
Db   2698 AAAAAA 2703

Query Match   3.5%;   Score 56.4;   DB 3;   Length 2719;
Best Local Similarity   90.9%;   Pred. No. 0.0028;
Matches   60;   Conservative   0;   Mismatches   6;   Indels   0;   Gaps   0;

RESULT 11
US-09-020-956-32/c
Sequence 32, Application US/09020956
Patent No. 6261562
GENERAL INFORMATION:
    APPLICANT: Xu, Jiangchun
    APPLICANT: Dillon, Davin C.
    TITLE OF INVENTION: COMPOUNDS FOR IMMUNOTHERAPY OF PROSTATE CANCER AND METHODS FO
    NUMBER OF SEQUENCES: 178
CORRESPONDENCE ADDRESS:
    ADDRESSEE: SEED and BERRY LLP
    STREET: 6300 Columbia Center, 701 Fifth Avenue
    CITY: Seattle
    STATE: WA
    COUNTRY: USA
    ZIP: 98104
COMPUTER READABLE FORM:
    MEDIUM TYPE: Floppy disk
    COMPUTER: IBM PC compatible
    OPERATING SYSTEM: PC-DOS/MS-DOS
    SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
    APPLICATION NUMBER: US/09/020,956
    FILING DATE: 09-FEB-1998

    CLASSIFICATION:
ATTORNEY/AGENT INFORMATION:
    NAME: Maki, David J.
    REGISTRATION NUMBER: 31,392
    REFERENCE/DOCKET NUMBER: 210121.427C2
TELECOMMUNICATION INFORMATION:
    TELEPHONE: (206) 622-4900
    TELEFAX: (206) 682-6031
INFORMATION FOR SEQ ID NO: 32:
SEQUENCE CHARACTERISTICS:
    LENGTH: 789 base pairs
    TYPE: nucleic acid
    STRANDEDNESS: single
    TOPOLOGY: linear
MOLECULE TYPE: cDNA
US-09-020-956-32

Qy   1503 CCAGCCTCCCATCTGACTTACAGCCAGGGTGGGGACTCAGCGGGTATAAAGCTAGTGTGAC 1562
           || |||| | || | |||| || || ||| || ||| ||||||||||||||
Db   129 CCTGTTGCCGTCCTGTCCCCAGCCTGTGNTGTGTCCCGNGAGGTTGTCAATAAAACCTGCC   70

Query Match   3.5%;   Score 55.8;   DB 4;   Length 789;
Best Local Similarity   67.6%;   Pred. No. 0.0026;
Matches   75;   Conservative   0;   Mismatches   36;   Indels   0;   Gaps   0;

RESULT 12
US-09-030-607-32/c
Sequence 32, Application US/09030607
Patent No. 6262245
GENERAL INFORMATION:
    APPLICANT: Xu, Jiangchun
    APPLICANT: Dillon, Davin C.
    TITLE OF INVENTION: COMPOUNDS FOR IMMUNOTHERAPY OF PROSTATE CANCER AND METHODS
    NUMBER OF SEQUENCES: 224
CORRESPONDENCE ADDRESS:
    ADDRESSEE: SEED and BERRY LLP
    STREET: 6300 Columbia Center, 701 Fifth Avenue
    CITY: Seattle
    STATE: WA
    COUNTRY: USA
    ZIP: 98104
COMPUTER READABLE FORM:
    MEDIUM TYPE: Floppy disk
    COMPUTER: IBM PC compatible
    OPERATING SYSTEM: PC-DOS/MS-DOS
    SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
    APPLICATION NUMBER: US/09/030,607
    FILING DATE: 25-FEB-1998
    CLASSIFICATION:
ATTORNEY/AGENT INFORMATION:
    NAME: Maki, David J.
    REGISTRATION NUMBER: 31,392
    REFERENCE/DOCKET NUMBER: 210121.427C3
TELECOMMUNICATION INFORMATION:
    TELEPHONE: (206) 622-4900
    TELEFAX: (206) 682-6031
INFORMATION FOR SEQ ID NO: 32:
SEQUENCE CHARACTERISTICS:
    LENGTH: 789 base pairs
    TYPE: nucleic acid
    STRANDEDNESS: single
    TOPOLOGY: linear
MOLECULE TYPE: cDNA
US-09-030-607-32

Qy   1563 CCTGAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 1613
           | ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   69 TNGGNAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA   19

Query Match   3.5%;   Score 55.8;   DB 4;   Length 789;
Best Local Similarity   67.6%;   Pred. No. 0.0026;

Matches 75; Conservative 0; Mismatches 36; Indels 0; Gaps 0;

RESULT 13
US-09-605-785-32/c
Sequence 32, Application US/0960S785
Patent No. 6321716
GENERAL INFORMATION:
APPLICANT: Xu, Jiangchun
APPLICANT: Dillon, Davin C.
APPLICANT: Mitcham, Jennifer L.
APPLICANT: Harlocker, Susan L.
APPLICANT: Jiang, Yuqui
APPLICANT: Henderson, Robert A.
APPLICANT: Kalos, Michael D.
APPLICANT: Fanger, Gary R.
APPLICANT: Retter, Marc W.
APPLICANT: Stolk, John A.
APPLICANT: Day, Craig H.
APPLICANT: Vedvick, Thomas S.
APPLICANT: Carter, Darrick
APPLICANT: Li, Samuel
APPLICANT: Wang, Aijun
APPLICANT: Skeiky, Yasir A.W.
APPLICANT: Hepler, William
TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY AND
TITLE OF INVENTION: DIAGNOSIS OF PROSTATE CANCER
FILE REFERENCE: 210121.427C16
CURRENT APPLICATION NUMBER: US/09,605,785
CURRENT FILING DATE: 2000-06-27
NUMBER OF SEQ ID NOS: 835
SOFTWARE: FastSEQ for Windows Version 3.0

SEQ ID NO 32
LENGTH: 789
TYPE: DNA
ORGANISM: Homo sapien
FEATURE:
NAME/KEY: misc_feature
LOCATION: (1)...(789)
OTHER INFORMATION: n = A,T,C or G
US-09-605-785-32

Query Match   3.5%;   Score 55.8;   DB 4;   Length 789;
Best Local Similarity 67.6%;   Pred. No. 0.0026;
Matches 75; Conservative 0; Mismatches 36; Indels 0; Gaps 0;

Qy 1503 CCAGCCTCCCATCTGACTACAGCCAGGGTGGGGACTCAGCGGGGTATAAAGCTAGTGTGAC 1562
          |||| ||||||| || | ||||||| ||| ||| ||| ||||||||||||||||
Db 129 CCTGTTGCCGTCCTGTCCCCAGCCTGNTTGTGTCCCGNGAGGTTGTCAATAAAACCTGCCC 70

Qy 1563 CCTGAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 1613
          | ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   69 TNGGNAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 19

RESULT 14
US-09-439-313-32/c
Sequence 32, Application US/0943S313
Patent No. 6329505
GENERAL INFORMATION:
APPLICANT: Xu, Jiangchun
APPLICANT: Dillon, Davin C.
APPLICANT: Mitcham, Jennifer L.
APPLICANT: Harlocker, Susan Louise
APPLICANT: Jiang Yuqui

APPLICANT: Reed, Steven G.
APPLICANT: Kalos, Michael
APPLICANT: Fanger, Gary
APPLICANT: Retter, Mark
APPLICANT: Solk, John
APPLICANT: Day, Craig
TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THERAPY AND
TITLE OF INVENTION: DIAGNOSIS OF PROSTATE CANCER
FILE REFERENCE: 210121.427C9
CURRENT APPLICATION NUMBER: US/09,439,313
CURRENT FILING DATE: 1999-11-12
NUMBER OF SEQ ID NOS: 575
SOFTWARE: FastSEQ for Windows Version 3.0

SEQ ID NO 32
LENGTH: 789
TYPE: DNA
ORGANISM: Homo sapien
FEATURE:
NAME/KEY: misc_feature
LOCATION: (1)...(789)
OTHER INFORMATION: n = A,T,C or G
US-09-439-313-32

Query Match   3.5%;   Score 55.8;   DB 4;   Length 789;
Best Local Similarity 67.6%;   Pred. No. 0.0026;
Matches 75; Conservative 0; Mismatches 36; Indels 0; Gaps 0;

Qy 1503 CCAGCCTCCCATCTGACTACAGCCAGGGTGGGGACTCAGCGGGGTATAAAGCTAGTGTGAC 1562
          |||| ||||||| || | ||||||| ||| ||| ||| ||||||||||||||||
Db 129 CCTGTTGCCGTCCTGTCCCCAGCCTGNTTGTGTCCCGNGAGGTTGTCAATAAAACCTGCCC 70

Qy 1563 CCTGAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 1613
          | ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   69 TNGGNAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 19

RESULT 15
US-09-352-616A-32/c
Sequence 32, Application US/0935S616A
Patent No. 6395278
GENERAL INFORMATION:
APPLICANT: Dillon, Davin C.
APPLICANT: Harlocker, Susan Louise
APPLICANT: Jiang, Yuqui
APPLICANT: Xu, Jiangchun
APPLICANT: Mitcham, Jennifer Lynn
TITLE OF INVENTION: COMPOUNDS FOR IMMUNOTHERAPY AND DIAGNOSIS
TITLE OF INVENTION: OF PROSTATE CANCER AND METHODS FOR THEIR USE
FILE REFERENCE: 210121.427C8
CURRENT APPLICATION NUMBER: US/09,352,616A
CURRENT FILING DATE: 1999-07-13
NUMBER OF SEQ ID NOS: 472
SOFTWARE: FastSEQ for Windows Version 3.0

SEQ ID NO 32
LENGTH: 789
TYPE: DNA
ORGANISM: Homo sapien
FEATURE:
NAME/KEY: misc_feature
LOCATION: (1)...(789)
OTHER INFORMATION: n = A,T,C or G
US-09-352-616A-32

Query Match   3.5%;   Score 55.8;   DB 4;   Length 789;
Best Local Similarity 67.6%;   Pred. No. 0.0026;
Matches 75; Conservative 0; Mismatches 36; Indels 0; Gaps 0;

Qy 1503 CCAGCCTCCCATCTGACTACAGCCAGGGTGGGGACTCAGCGGGGTATAAAGCTAGTGTGAC 1562
          |||| ||||||| || | ||||||| ||| ||| ||| ||||||||||||||||
Db 129 CCTGTTGCCGTCCTGTCCCCAGCCTGNTTGTGTCCCGNGAGGTTGTCAATAAAACCTGCCC 70

Qy 1563 CCTGAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 1613
          | ||||||||||||||||||||||||||||||||||||||||||||||||||||||||

Db   69 TNGGNAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 19

Search completed: March 29, 2003, 22:23:46
Job time : 105.085 secs

THIS PAGE BLANK (USPTO)

```
                              GenCore version 5.1.4_p5_4578
                              Copyright (c) 1993 - 2003  Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on:           March 29, 2003, 21:53:16 ; Search time 170.281 seconds
                                              (without alignments)
                                              8080.893 Million cell updates/sec

Title:            US-09-691-219-1
Perfect score:    1617
Sequence:         1 cgcggtgaacagcacgaggg..........aaaaaaaaaaaaaatgtt 1617

Scoring table:    IDENTITY_NUC
                  Gapop 10.0 , Gapext 1.0

Searched:         574371 seqs, 425486471 residues

Total number of hits satisfying chosen parameters:       1148742

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing:  Minimum Match 0%
                  Maximum Match 100%
                  Listing first 45 summaries

Database :        Published_Applications_NA.*
                  1:/cgn2_6/ptodata/2/pubpna/US07_PUBCOMB.seq:*
                  2:/cgn2_6/ptodata/2/pubpna/PCT_NEW_PUB.seq:*
                  3:/cgn2_6/ptodata/2/pubpna/PCT_NEW_PUB.seq:*
                  4:/cgn2_6/ptodata/2/pubpna/US06_NEW_PUB.seq:*
                  5:/cgn2_6/ptodata/2/pubpna/US06_PUBCOMB.seq:*
                  6:/cgn2_6/ptodata/2/pubpna/PCTUS_PUBCOMB.seq:*
                  7:/cgn2_6/ptodata/2/pubpna/US08_NEW_PUB.seq:*
                  8:/cgn2_6/ptodata/2/pubpna/US08_PUBCOMB.seq:*
                  9:/cgn2_6/ptodata/2/pubpna/US09_NEW_PUB.seq:*
                 10:/cgn2_6/ptodata/2/pubpna/US09_PUBCOMB.seq:*
                 11:/cgn2_6/ptodata/2/pubpna/US10_NEW_PUB.seq:*
                 12:/cgn2_6/ptodata/2/pubpna/US10_PUBCOMB.seq:*
                 13:/cgn2_6/ptodata/2/pubpna/US60_NEW_PUB.seq:*
                 14:/cgn2_6/ptodata/2/pubpna/US60_PUBCOMB.seq:*
```

SUMMARIES

```
                                  %
Result              Score   Query    Match Length DB ID            Description
No.                         Match
-------             -----   -----   -------------- -----           -----------
  1     1590.2   98.3    1630    9   US-10-162-012-39    Sequence 39, Appl
  2     1117.4   69.1    1119    9   US-10-162-012-41    Sequence 41, Appl
  3      309.6   19.1    2852   10   US-09-872-153-20    Sequence 20, Appl
  4      175.6   10.9    1826    9   US-09-957-708-13    Sequence 13, Appl
  5       73.8    4.6     278   10   US-09-294-093B-522  Sequence 522, App
  6       61.8    3.8     900   10   US-09-982-809-21    Sequence 21, Appl
  7       61      3.8    1254    9   US-09-986-480-19    Sequence 19, Appl
  8       58.6    3.6    2378   10   US-10-033-350-1     Sequence 1, Appl
  9       58.2    3.6     307    9   US-10-091-483-54    Sequence 54, Appl
 10       58.2    3.6     307   10   US-09-764-846-54    Sequence 54, Appl
 11       58.2    3.6     314   10   US-10-091-483-126   Sequence 126, App
 12       58.2    3.6     314    9   US-09-764-846-126   Sequence 126, Appl
 13       57.4    3.5     371    9   US-09-925-299-16    Sequence 16, Appl
 14       57.4    3.5     371   10   US-09-764-299-16    Sequence 16, Appl
 15       57      3.5    1152   10   US-09-745-763-149   Sequence 149, App
 16       57      3.5    2269   10   US-09-925-299-1     Sequence 1, Appl
 17       56.6    3.5    2536   10   US-09-767-870-2     Sequence 2, Appli
 18       56      3.5     552    9   US-10-012-542-111   Sequence 111, App
 19       56      3.5    1296    9   US-10-012-542-29    Sequence 29, Appl
```

```
c   20   55.8   3.5    789    9   US-09-232-880-32    Sequence 32, Appl
c   21   55.8   3.5    789    9   US-10-012-896-32    Sequence 32, Appl
c   22   55.8   3.5    789    9   US-09-895-793-32    Sequence 32, Appl
c   23   55.8   3.5    789    9   US-09-895-814-32    Sequence 32, Appl
c   24   55.8   3.5    789   10   US-09-759-143-32    Sequence 32, Appl
c   25   55.8   3.5    789   10   US-09-780-669-32    Sequence 32, Appl
c   26   55.8   3.5    789   10   US-09-030-606-32    Sequence 32, Appl
c   27   55.8   3.5    789   10   US-09-822-687-32    Sequence 32, Appl
c   28   55.8   3.5    789   10   US-09-115-453-32    Sequence 32, Appl
c   29   55.8   3.4    321   10   US-09-960-352-6165  Sequence 6165, Ap
c   30   55.6   3.4   1063    9   US-10-102-806-231   Sequence 231, App
c   31   55.6   3.4   1331    9   US-09-764-884-11    Sequence 11, Appl
c   32   55.6   3.4   1331   10   US-09-745-605-11    Sequence 11, Appl
c   33   55.6   3.4   1408   10   US-09-092-256-11    Sequence 3, Appl1
c   34   55.6   3.4   1790    9   US-09-944-413-31    Sequence 31, Appl
c   35   55.6   3.4   1790    9   US-09-944-403-31    Sequence 31, Appl
c   36   55.6   3.4   1790    9   US-09-944-896-31    Sequence 31, Appl
c   37   55.6   3.4   1790    9   US-09-944-944-31    Sequence 31, Appl
c   38   55.6   3.4   1790    9   US-09-944-907-31    Sequence 31, Appl
c   39   55.6   3.4   1790    9   US-09-944-029-31    Sequence 31, Appl
c   40   55.6   3.4   1790   10   US-09-866-028-31    Sequence 31, Appl
c   41   55.6   3.4   1790   10   US-09-944-449-31    Sequence 31, Appl
c   42   55.6   3.4   1790   10   US-09-944-457-31    Sequence 31, Appl
c   43   55.6   3.4   1790   10   US-09-945-587-31    Sequence 31, Appl
c   44   55.6   3.4   1790   10   US-09-945-015-31    Sequence 31, Appl
c   45   55.6   3.4   1790   10   US-09-944-396-31    Sequence 31, Appl
```

ALIGNMENTS

```
RESULT 1
US-10-162-012-39
; Sequence 39, Application US/10162012
; Publication No. US20030051660A1
; GENERAL INFORMATION:
; APPLICANT:  Curtis, Rory A.J.
; APPLICANT:  Silos-Santiago, Inmaculada
; APPLICANT:  Gu, Wei
; TITLE OF INVENTION: NOVEL HUMAN ION CHANNEL AND TRANSPORTER FAMILY MEMBERS
; FILE REFERENCE: 10448-190001
; CURRENT APPLICATION NUMBER: US/10/162,012
; CURRENT FILING DATE:  2002-06-04
; PRIOR APPLICATION NUMBER: US 60/209,845
; PRIOR FILING DATE: 2000-06-06
; PRIOR APPLICATION NUMBER: US 09/875,321
; PRIOR FILING DATE: 2001-06-06
; PRIOR APPLICATION NUMBER: US 09/875,423
; PRIOR FILING DATE: 2001-06-05
; PRIOR APPLICATION NUMBER: US 60/209,257
; PRIOR FILING DATE: 2000-06-05
; PRIOR APPLICATION NUMBER: US 09/875,423
; PRIOR FILING DATE: 2001-06-05
; PRIOR APPLICATION NUMBER: PCT/US01/18340
; PRIOR FILING DATE: 2001-06-06
; PRIOR APPLICATION NUMBER: PCT/US01/18398
; PRIOR FILING DATE: 2001-06-05
; PRIOR APPLICATION NUMBER: US 60/209,238
; PRIOR FILING DATE: 2000-06-05
; PRIOR APPLICATION NUMBER: US 09/875,363
; PRIOR FILING DATE: 2001-06-05
; PRIOR APPLICATION NUMBER: US 09/18247
; PRIOR FILING DATE: 2001-06-05
; PRIOR APPLICATION NUMBER: PCT/US01/18247
; PRIOR FILING DATE: 2001-06-05
; PRIOR APPLICATION NUMBER: US 60/227,068
; PRIOR FILING DATE: 2000-08-22
; PRIOR APPLICATION NUMBER: US 60/227,068
; PRIOR FILING DATE: 2000-08-22
; PRIOR APPLICATION NUMBER: US 09/928,530
; PRIOR FILING DATE: 2001-08-13
; PRIOR APPLICATION NUMBER: PCT/US01/25475
; PRIOR FILING DATE: 2001-08-15
; PRIOR APPLICATION NUMBER: US 60/226,770
; PRIOR FILING DATE: 2000-08-21
; PRIOR APPLICATION NUMBER: US 09/934,421
; PRIOR FILING DATE: 2001-08-21
; PRIOR APPLICATION NUMBER: PCT/US01/26096
; PRIOR FILING DATE: 2001-08-21
```

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

```
PRIOR APPLICATION NUMBER: US 60/279,281
PRIOR FILING DATE: 2001-03-28
PRIOR APPLICATION NUMBER: US 10/109,029
PRIOR FILING DATE: 2002-03-28
PRIOR APPLICATION NUMBER: PCT/US02/09728
PRIOR FILING DATE: 2002-03-28
PRIOR APPLICATION NUMBER: US 60/290,288
PRIOR FILING DATE: 2001-05-11
PRIOR APPLICATION NUMBER: US (not assigned)
PRIOR FILING DATE: 2001-05-13
NUMBER OF SEQ ID NOS: 48
SOFTWARE: FastSEQ for Windows Version 4.0

SEQ ID NO 39
LENGTH: 1630
TYPE: DNA
ORGANISM: Homo sapiens
FEATURE:
NAME/KEY: CDS
LOCATION: (230)...(1345)
US-10-162-012-39

Query Match       98.3%;  Score 1590.2;  DB 9;  Length 1630;
Best Local Similarity  99.8%;  Pred. No. 0;  Mismatches 0;
Matches 1592;  Conservative  0;  Mismatches  3;  Indels  0;  Gaps  0;
```

```
Qy    20 GGGAGACGCGCTTGGCACGCGGGACCCAGAGACCCGAGCCTCCAGCCCGCACGGGACCCGGGGGGAG  79
         ||||||||| |||||||||||||||||||||||||||||||||||||||||| ||||||||||||||
Db    20 GGGAGACGCTTGGCACGCGGGACCCAGAGACCCGAGCCTGCAGCCCGCACGGGGACCCGGGGGGAG   79

Qy    80 CTCAGGAGGGGCCGCAGGAGCCGGGGCCCGGAGTGAGCGCACCTCGCGGGGCCCTCGGGGCCAGGT  139
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    80 CTCAGGGGCCGCAGGAGCCGGGGCCCGGAGTGAGCGCACCTCGCGGGGCCCTCGGGGCCAGGT    139

Qy   140 GGGTGAGCGGCGCCACCCGGGAGTCCCCGGGGCGCCAACTTTCAGGCGCGCCACTCGGGCGGCTG  199
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   140 GGGTGAGCGGCGCCACCCGGGAGTCCCCGGGGCGCCAACTTTCAGGCGCGCCACTCGGGCGGCTG  199

Qy   200 CGGCGGCTGCCCGGGGACTCGCGGCGCGGGACTGCATGGAGGGCCAAGGAGAAGCAGCAGCATCTGTTG  259
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   200 CGGCGGCTGCCCGGGGACTCGCGGCGCGGGACTGCATGGAGGGCCAAGGAGAAGCAGCCATCTGTTG  259

Qy   260 GACGCCGCAGGCGCGGCAAATCCGGTCATACACGGGGATCTCTGTGGGCGAGAAAGGGGGCTGGCTGG  319
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   260 GACACCGCAGGCGCGGCAATCCGGTCATACACGGGGATCTCTGTGGGCGAAAAGGGGGCTGGCTGG  319

Qy   320 ATTCCTCTCGCCCCGACCTGCCTGACTTGCCAGGCGCCATTGAGCTGCCTGGCTGC  379
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   320 ATTCCTCTCGCCCCGACCTGCCTGACTTGCCAGGCGCCATTGAGCTGCCTGGCTGC  379

Qy   380 CATCACTGCCATGCTCAGAGAAGGGTCCTGACCAGTCACTGTGACCCCAAGAACGGGGAAGGCC  439
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   380 CATCACTGCCATGCTCAGAGAAGGGTCCTGACCAGTCACTGTGACCCCAAGAACGGGGAAGGCC  439

Qy   440 CAGCGCCCAGCGTGTAGCCTCTGCCATCCTGCCGTCCTGTTGTTCATCGGAGAAGTCGTT  499
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   440 CAGCGCCCAGCGTGTATTAGCCTCTGCCATCTGCCTGTTGTTCATCGGAGAAGTCGTT  499

Qy   500 GGTGGGGTACCTGGCACACAGCTTGGCTGTCATGAGGACTGACGCACACCTGCTCACTGAC  559
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   500 GGTGGGGTACCTGGCACACAGCTTGGCTGTCATGAGGACTGACGCACACCTGCTCACTGAC  559

Qy   560 TTTGCCAGCATGGCTCATCAGCTCTTCTCTCCCTCTCGATGTGTCCGGGCCACCAAG  619
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   560 TTTGCCAGCATGGCTCATCAGCCTCTTCTCCCTCTCGGATGTGTCCGGGCCACCAAG  619

Qy   620 ACCATGAACTTTGGGCTGGCAGAGAGACTTGAGATCTTGGGAGCCCTGGTCTCTCTACTGTCC  679
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   620 ACCATGAACTTTGGGCTGGCAGAGACTTGAGATCTTGGGAGCCCTGGTCTCTCTACTGTCC  679

Qy   680 ATCTGGGTCGTGACGGGGGTGGTACTGGTTGTACCTGGGAC  739
         |||||||||||||||||||||||||||||||||||||||||
Db   680 ATCTGGGTCGTGACGGGGGTGGTACTGGTTGTACCTGGGAC  739

Qy   740 TATGAAAATTGACGGGGGGGACCCATGCTGANTCACGTCGGCGCTGGCGTGTGAACATC  799
         |||||| |||||||||||||||||||||||| ||||||||||||||||||||||||||
Db   740 TATGAAATTGACGGGGGGGACCCATGCTGATCACGTCGGCGCTGGCGTGTGAACATC   799

Qy   800 ATAATGGGGTTGACCCTTCACCAGTCTGGCCATGGCACAGCCCACCACCAACCAG  859
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   800 ATAATGGGGTTGACCCTTCACCAGTCTGGCCATGGCACAGCCCACCACCAACCAG  859

Qy   860 CAGGAGGAGAACCCCAGCGTCCGAGCTGCCTTCATCATGGATGATCGGCGACTTTATCCAG  919
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   860 CAGGAGGAGAACCCCAGCGTCCGAGCTGCCTTCATCATGGATGATCGGCGACTTTATCCAG  919

Qy   920 AGCATGGGGTGTCCTTAGTGGCAGCCTATATTTTATACTTCAAGCCAGAATACAAGTATGTA  979
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   920 AGCATGGGGTGTCCTTAGTGGCAGCCTATATTTTATACTTCAAGCCAGAATACAAGTATGTA  979

Qy   980 GACCCCCATCTGCACCTTTGTCTTCTCCAATCCTGGTCCTGGGGGACAAACCTTGACCATCCTG 1039
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   980 GACCCCCATCTGCACCTTTGTCTTCTCCAATCCTGGTCCTGGGGGACAAACCTTGACCATCCTG 1039

Qy  1040 AGAGATGTGATCATCCTGGTGTTGATGGGAAGGGGACCCCCAAGGGCGGTTGACTTCACGTGTT 1099
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1040 AGAGATGTGATCATCCTGGTGTTGATGGGAAGGGGACCCCCAAGGGCGGTTGACTTCACAGCTGTT 1099

Qy  1100 CGTGATCTGCTGTTGTCGGTGGAGGGGTAGAAGCCCTGCACAGCCTGCATATGTCATATCTGGGCA 1159
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1100 CGTGATCTGCTGTTGTCGGTGGAGGGGTAGAAGCCCTGCACAGCCTGCATATCTGGGCA 1159

Qy  1160 CTGACGGTGGCCCCAGCCTTGTTCTGTCTCACATCGCCATTGCTTCAGAATACAGAGACCC 1219
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1160 CTGACGGTGGCCCCAGCCTTGTTCTGTCTCACATCGCCATTGCTTCAGAATACAGAGACCC 1219

Qy  1220 CAGGCTGTTGCTGAAGACAGCCAGCAGCCTCCAAGGGAGAGTTCCACTTCACACCGTG 1279
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1220 CAGGCTGTTGCTGAAGACAGCCAGCAGCCTCCAAGGGAGAGTTCCACTTCACACCGTG 1279

Qy  1280 ACCATCCAGATCGAGGACTACTCGGAGGAGGACAATGAAGGACCCTGCACAGCCTCCATATCTGGGCA 1339
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1280 ACCATCCAGATCGAGGACTACTCGGAGGAGGACAATGAAGGACCCTGCACAGCCTCCATATCTGGGCA 1339

Qy  1340 TCAGACTGACTGCTCAGCCGAGGCACCAACTGGGCATGGGGCCATGGGTCAGGTGGCTGG 1399
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1340 TCAGACTGACTGCTCAGCCGAGGCACCAACTGGGCATGGGGCCATGGGTCAGGTGGCTGG 1399

Qy  1400 ACTGAGTGTCCCCCAGGCCCCAGCCGAGGCCATGTGCTTGTTATAAAACCAGG 1459
         ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1400 ACTGAGTGTCCCCCAGGCCCCAGCCGAGGCCATGTGCTTGTTATAAAACCAGG 1459

Qy  1460 TCCCCCTCCTGACCTCTGCCCCACTCCAGGAATGGGAGTCTTCCCAGCTCCATCTGAC 1519
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1460 TCCCCCTCCTGACCTCTGCCCCACTCCAGGAATGGGAGTCTTCCCAGCTCCATCTGAC 1519

Qy  1520 TACAGGCCAGGGTGGGGGGACTCAGGGGTATAAAGCTAGTGTGACCCTGAAAAAAAAAA 1579
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1520 TACAGCGCCAGGGTGGGGGGTATAAAGCTAGTGTGACCCTGAAAAAAAAAA 1579

Qy  1580 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAT 1614
         |||||||||||||||||||||||||||||
Db  1580 AAAAAAAAAAAAAAAAAAAAAAAAAAAGCAT 1614
```

```
;   PRIOR APPLICATION NUMBER: US 60/209,845
;   PRIOR FILING DATE: 2000-06-06
;   PRIOR APPLICATION NUMBER: US 09/875,321
;   PRIOR FILING DATE: 2001-06-06
;   PRIOR APPLICATION NUMBER: PCT/US01/18340
;   PRIOR FILING DATE: 2001-06-06
;   PRIOR APPLICATION NUMBER: US 60/209,257
;   PRIOR FILING DATE: 2000-06-05
;   PRIOR APPLICATION NUMBER: US 09/875,423
;   PRIOR FILING DATE: 2001-06-05
;   PRIOR APPLICATION NUMBER: PCT/US01/18398
;   PRIOR FILING DATE: 2001-06-05
;   PRIOR APPLICATION NUMBER: US 60/209,238
;   PRIOR FILING DATE: 2000-06-05
;   PRIOR APPLICATION NUMBER: US 09/875,363
;   PRIOR FILING DATE: 2001-06-05
;   PRIOR APPLICATION NUMBER: PCT/US01/18247
;   PRIOR FILING DATE: 2001-06-05
;   PRIOR APPLICATION NUMBER: US 60/227,068
;   PRIOR FILING DATE: 2000-08-22
;   PRIOR APPLICATION NUMBER: US 09/928,530
;   PRIOR FILING DATE: 2001-08-13
;   PRIOR APPLICATION NUMBER: PCT/US01/25475
;   PRIOR FILING DATE: 2001-08-15
;   PRIOR APPLICATION NUMBER: US 60/226,770
;   PRIOR FILING DATE: 2000-08-21
;   PRIOR APPLICATION NUMBER: US 09/934,421
;   PRIOR FILING DATE: 2001-08-21
;   PRIOR APPLICATION NUMBER: PCT/US01/26096
;   PRIOR FILING DATE: 2001-08-21
;   PRIOR APPLICATION NUMBER: US 60/279,281
;   PRIOR FILING DATE: 2001-03-28
;   PRIOR APPLICATION NUMBER: US 10/109,029
;   PRIOR FILING DATE: 2002-03-28
;   PRIOR APPLICATION NUMBER: PCT/US02/09728
;   PRIOR FILING DATE: 2002-03-28
;   PRIOR APPLICATION NUMBER: US 60/290,288
;   PRIOR FILING DATE: 2001-05-11
;   PRIOR APPLICATION NUMBER: US (not assigned)
;   PRIOR FILING DATE: 2002-05-13
;   NUMBER OF SEQ ID NOS: 48
;   SOFTWARE: FastSEQ for Windows Version 4.0
;
; SEQ ID NO 41
;   LENGTH: 1119
;   TYPE: DNA
;   ORGANISM: Homo sapiens
US-10-162-012-41

  Query Match             69.1%;
  Best Local Similarity   99.8%;
  Matches  1118;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

  Score 1117.4;  DB 9;  Length 1119;
  Pred. No. 1.1e-280;

Qy          230 ATGGAGGGCCAAGGAGAAGAAGCAGCAGCAATCTGTTGGACGGACGGCCGCCAATCCGGTCATACACG 289
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db            1 ATGGAGGGCCAAGGAGAAGAAGCAGCAGCAATCTGTTGGACGGACGGCCGCCAATCCGGTCATACACG 60

Db          301 ATGACTGACGCAGCACACCTGCTCACTGACTTTGCCAGCATGCTCATCAGCCTCTTCTCC 360

Qy          290 GGATCTCTGTGCGGAGGAAGGGCTGGCTGGATTCCTCTGCCCCCGACCTGGCCGACTTG 349
                |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db           61 GGATCTCTGTGCGGAGGAAGGGCTGGCTGGATTCCTCTGCCCCCGACCTGGCCGACTTG 120

Qy          350 CAGGCCAATGAGCTGGCTGGCCCAGAGCAAACCATCACTGCCATGCTCAGCAAGGGGTCCTGAC 409
                |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          121 CAGGCCAATGAGCTGGCTGGCCCAGAGCAAACCATCACTGCCATGCTCAGCAAGGGGTCCTGAC 180

Qy          410 AGTCACTGTGACCCCAGAAAGGGAAGGCCCAGCTGTGATGTAGCCTCTGCCATC 469
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          181 AGTCACTGTGACCCCAGAAAGGGAAGGCCCAGCTGTGATGTAGCCTCTGCCATC 240

Qy          470 TGCCTGTTGTTCATGATCGGAGAAGTCGTTGGTGGGTACCTTGGCCTGTC 529
                |||||||||||||||||||||||||||||||||||||||||||||||||||
Db          241 TGCCTGTTGTTCATGATCGGAGAAGTCGTTGGTGGGTACCTTGGCCTGTC 300

Qy          530 ATGACTGACGCAGCACACCTGCTCACTGACTTTGCCAGCATGCTCATCAGCCTCTTCTCC 589
```

```
Db          301 ATGACTGACGCAGCACACCTGCTCACTGACTTTGCCAGCATGCTCATCAGCCTCTTCTCC 360

Qy          590 CTCTGGATGTCCTCCCGGCCCAGCCACCACCAAGACCATGAAACTTTGGGTGTGGCAGGAGGACTGAG 649
                |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          361 CTCTGGATGTCCTCCCGGCCCAGCCACCACCAAGACCATGAAACTTTGGGTGTGGCAGGAGGACTGAG 420

Qy          650 ATCTTGGGACCCCTGTCTCTGTACTGTCCATCCTGGGGGTCCTGACCGGGGGGTACTGGTGTAC 709
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          421 ATCTTGGGACCCCTGTCTCTGTACTGTCCATCCTGGGGGTCCTGACCGGGGGGTACTGGTGTAC 480

Qy          710 CTGGCCTGTGGAGCGGCTGATCTCTGGGGACTATGAAAATTGACGGGGGGGGGGGGGACCATGCTGATC 769
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          481 CTGGCCTGTGGAGCGGCTGATCTCTGGGGACTATGAAAATTGACGGGGGGGGGGGGGACCATGCTGATC 540

Qy          770 ACGTCGGGGCGTGGCGCTGTGGGCGCTGTGGGTTGACCCTTCACCAGTCTGGC 829
                |||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          541 ACGTCGGGGCGTGGCGCTGTGGGCGCTGTGGGTTGACCCTTCACCAGTCTGGC 600

Qy          830 CATGGGCACAGCCACGCACCACCAGCAGGAGGAGAACCCAGCGTCCGACGTGCC 889
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          601 CATGGGCACAGCCACGCACCACCAGCAGGAGGAGAACCCAGCGTCCGACGTGCC 660

Qy          890 TTCATCCATGTGATGGGGCGACTTTATGCAGAGCATGGGTGTCCTAGTGGCAGCCTATATT 949
                |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          661 TTCATCCATGTGATGGGGCGACTTTATGCAGAGCATGGGTGTCCTAGTGGCAGCCTATATT 720

Qy          950 TTATACTTCAAGCCAGAAATACAAGTATGTGAGACCCCACCTTCGTCTTCCATC 1009
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          721 TTATACTTCAAGCCAGAAATACAAGTATGTGAGACCCCACCTTCGTCTTCCATC 780

Qy         1010 CTGGTCCTGGGGACAACCTTGACCATCCTGAGAAGATGTGATCCTGGTGGTTGTTGAATGG 1069
                |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          781 CTGGTCCTGGGGACAACCTTGACCATCCTGAGAAGATGTGATCCTGGTGGTTGTTGAATGG 840

Qy         1070 ACCCCCAAGGGCGTTGACTTCACAGCTGTTCGTGATCGTCGTGGTTGGAGGGGGGTA 1129
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          841 ACCCCCAAGGGCGTTGACTTCACAGCTGTTCGTGATCGTCGTGGTTGGAGGGGGGTA 900

Qy         1130 GAAGCCCTGCACACAGCCTGCATATCTGGGCACTGACGGTGGCCCAGCCTGTTCTGTTCGTC 1189
                |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          901 GAAGCCCTGCACACAGCCTGCATATCTGGGCACTGACGGTGGCCCAGCCTGTTCTGTTCGTC 960

Qy         1190 CACATCGCCATTGCTCAGAATACAGACGCCCAGGGTGTTGCTGAAGACAGCCAGCAGCCCGC 1249
                |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          961 CACATCGCCATTGCTCAGAATACAGACGCCCAGGGTGTTGCTGAAGACAGCCAGCAGCCCGC 1020

Qy         1250 CTCCAAGGGAAGTTCCACTTCCACACGGTGACCATCCAGATCCGAGGACTACTCGGAGGAC 1309
                |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         1021 CTCCAAGGGAAGTTCCACTTCCACACGGTGACCATCCAGATCCGAGGACTACTCGGAGGAC 1080

Qy         1310 ATGAAGGACTGTCAGGCATGCCAGGGCCCCTCAGACTGA 1348
                |||||||||||||||||||||||||||||||||||||||
Db         1081 ATGAAGGACTGTCAGGCATGCCAGGGCCCCTCAGACTGA 1119
```

```
RESULT 3
US-09-872-153-20
; Sequence 20, Application US/09872153
; Patent No. US20020082207A1
; GENERAL INFORMATION:
;   APPLICANT: Hirst, Shannon K.
;   APPLICANT: Harlocker, Susan L.
;   APPLICANT: Dillon, Davin C.
;   APPLICANT: Kalos, Michael D.
;   TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY
;   TITLE OF INVENTION:  AND DIAGNOSIS OF PANCREATIC CANCER
;   FILE REFERENCE: 210121.531
;   CURRENT APPLICATION NUMBER: US/09/872,153
;   CURRENT FILING DATE: 2001-05-31
;   NUMBER OF SEQ ID NOS: 28
;   SOFTWARE: FastSEQ for Windows Version 4.0
;
; SEQ ID NO 20
```

```
;
; LENGTH: 2852
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-872-153-20

Query Match    19.1%;    Score 309.6;  DB 10;  Length 2852;
Best Local Similarity  57.9%;    Pred. No. 1.3e-70;
Matches  574;  Conservative  0;  Mismatches  409;  Indels  9;  Gaps  1;

Qy    359 GAGCTGGCTGCCCAGACGAACCAACCATCGCCATGCTCAGAAGGGTGCCTGACAGTCACTGT 418
          ||||||||          |                 |      |||||||
Db    439 GAGCTGGAGTCAGGAGGGCATNGTACCACTGCCCACAGCCCACCAGAGAAAGGGG 498

Qy    419 GACCCCCAAGAAAGGGGAAAGGCCCAGCGCCCAGCGTGTANGTAGCCTCTGCCATCTGCCTGTTG 478
          ||||                  |         |  | |      |||||||||||
Db    499 GCGAATGAGTACGCCTATGCCAAGTGGAAAACTCGTTCTCCTTCAGCAATATNGCTTCATT 558

Qy    479 TTCATGATGATCGGAGAGAAGTCGTTGGTGGGTACCTGCGCACACAGCTTGGCTGTCATGACTGAC 538
          ||||||          |   |   |       |||||  |||||||
Db    559 TTTCATGATTCCAGAGGTCGTGGGGTGGGCACATTGCTGGGAGTCTTCGTCTGTTGTTCACAGAT 618

Qy    539 GCAGCACACCTGCTCACTGACTTTGCCAGCATGCTCATCAGCCTGTTCTCCTTCTCCCTCTGGATG 598
          |||||||||  |  ||  ||||   ||    |  |    ||||||||||||
Db    619 GCTGCCCACCTCTTRATTGACCCAGTTTCCTCGTCAGTCTCTTCTCCCTGTGGTTG 678

Qy    599 TCCTCCCGGCCCAGACCACCAGAACTTTGCCTGGCAGAGACTGAGATCTTGGGA 658
          |||||          |           |  ||  ||||||||||||
Db    679 TCATCGAAGCCTCCCTCTAAGCGGCTGACATTTGGATNGGCACCGAGCCAGAGAATCCTTGGT 738

Qy    659 GCCCTGGTCTCTGTACTGTCCCATCTGGGTCGTGACGGGGGTNACTGGTTGTTACCTGCTGTG 718
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    739 GCCCTGCTCTCCATCCTGTCCCATCTGGGTCGTGACTGCTTGTGTATGTCACCTGGCCATGT 798

Qy    719 GAGGGGGCTGATCTCGGGGGACTATGAAATTGACGCGGGGGGGGAAACATCACGTCGGGC 778
          ||||||||||||||||||||||||||||||||||||||||||||||||
Db    799 GAGCCGCCTCTGCTGTATCCTGTGATTACCAGGGGCACTGTGATGANTCACGTTTCCAGC 858

Qy    779 TGCGCGTGCTGGGCTGTGAACATCATNAATNGGGGTTGACCCTTCACCAGTCTCGGCCCATGGGCAC 838
          |||||||      |      ||||  |||||||||||||||||||||
Db    859 TGCGCAGTGGCGGCCAACATTGTTACTNAACGTGTGGTTTTGCACCCAGAGATGGCCTTGGCCAC 918

Qy    839 AGCCACGGCACCCACAACCAGCAGGAGGAGAAACCCCAGACGCTGCCTTCATCCAT 898
          ||||||||||||||||||||||||||||||||||||
Db    919 AATCA-------CAAGGAAGTACAAGCCAATGCCAGCGTCAGAGCTGCTTTTGTGCAT 969

Qy    899 GTGATCGGCGCGACTTTATGCAGAGCANGGGTGTCCTAGTGCGGCAGCCTAATATTTTTATACTTC 958
          ||||||||       |        |   ||||||||||||||||
Db    970 GCCCTTGGAGANTCTATTTCAGAGTAGTCGCGAGAGAAATTGCTAAAGCCCTTAGCAAA 1029

Qy    959 AAGCCAGAATACAACAAGTATAGTTAGAGACCCCCATCCTTCGTCTCTTCCATCCTGGTCCTG 1018
          ||||||||||          |     | |             ||||||||||||
Db   1029 GCCCTTGGAGATTCANTTTTTTAGCAGTCGACAGGGTGCCTGTCTGTG 1029

Qy   1019 GGGACAAACCTTTGACCATCCTGAGAGANTGGATGATCCTGGGTGTTGATGGAAAGGGACCCCCAAG 1078
          ||||||||          ||||||||||||||||||||||||||||
Db   1090 GCCAGCACCATCACATTCTTAAAGGACTTCTCCATCTTACTCACTGGGAAGGTGTGCCAAAG 1149

Qy   1079 GGCCGTTGACTTCACACGTCTTGGATCTGCGTCCGTGTGGAGGGGGTAGAAGCCCTG 1138
          ||||||||          ||  ||  |||||||||||||||||||||
Db   1150 AGCCTGAATTACAGTGCTGTGTGAAAGAGCTTATTTTTAGCAGTCGACAGGGTGCTGCTGTG 1209

Qy   1139 CACAGCCTGCATATCTGGGCACTGACGGGTGGCCCAGCCTGTTCTGTTGTCCATCGCC 1198
          ||||||||||||          |  ||  ||||||||||||
Db   1210 CACAGCCTGCACACTCGGGTCTCTAACAAATGAATCAAGTAATTCTTCAGCTCATGTTTGCT 1269

Qy   1199 ATTGCTCAGAATACAGACGCGCCAGCGACGCAGCACGCCGCGCGCCTTCAAGGG 1258
          |||||||          |   |  ||  |||||||||||||||||
Db   1270 ACAGGACCAGGCACGGGGACAGCCAAGTAGGTTCGGAGAGAAATTGCTAAAGCCCTTAGCAAA 1329

Qy   1259 AAGTTCCACTTCCACACCGTGACCATCCAGANTCGNAGGGACTACTCGGAGGACCGATGAGGAC 1318
          |||          ||      |  |   ||||||||||||||
Db   1330 AGCTTTACGANGCACTCACCAATTCAGANTGGAATNCTCCAGTTGACCAGGACCCGAC 1389

Qy   1319 TGTCAGGCCANGCCAGGCGCCCCTTCAGAGACTGACT 1350
          |          ||||||||||||||||||||||||||
Db   1390 TGCCCTTTTCTGTGAAGAGACCCCTGTGACTAGCT 1421
```

```
RESULT 4
US-09-957-708-13
; Sequence 13, Application US/09957708
; Publication No. US20030031678A1
GENERAL INFORMATION:
; APPLICANT: Sun, Yongming
; APPLICANT: Recipon, Herve
; APPLICANT: Gafferkey, Robert
; APPLICANT: Ali, Shujath
; TITLE OF INVENTION: Compositions and Methods Relating to Prostate Specific
; TITLE OF INVENTION: Genes
; FILE REFERENCE: DEX-0239
; CURRENT APPLICATION NUMBER: US/09/957,708
; CURRENT FILING DATE: 2001-09-19
; PRIOR APPLICATION NUMBER: 60/233,746
; PRIOR FILING DATE: 2000-09-19
; NUMBER OF SEQ ID NOS: 40
; SOFTWARE: Patentin Ver. 2.1
;
; SEQ ID NO 13
; LENGTH: 1826
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-957-708-13

Query Match    10.9%;    Score 175.6;  DB 9;  Length 1826;
Best Local Similarity  53.1%;    Pred. No. 7.2e-36;
Matches  431;  Conservative  0;  Mismatches  339;  Indels  42;  Gaps  1;

Qy    425 AAGAAAGGGGAAAGGCCCAGCGCCCAGCGTGTATGTAGCCTCTGCCATCTGCCTGTTGTTCATG 484
          ||||||||          |  |                 |   |      ||||||||
Db    585 AAGCAGAGAAAAGGTGAAAGCCAGGTGACCAGTTGCCGCGTTCTGTACTTGCTTTTCATG 644

Qy    485 ATCGGAGAAGTCGTTGGTGGGTACCTGCGCACACAGCTTGGCGTGTCATGACTGACGCAGCA 544
          ||||||||          |   |  ||||||||||||||
Db    645 ATTGGAGAACTTGTAGGTGGGATCANCATTGCAAAATAGCCCTAGCAANTCATGACAGATGCACTT 704

Qy    545 CACCTGTCACTGACTTTGCCAGCATGCTCATCAGCCCTCTTCTCCCCTCGGGATGGTCCTCC 604
          ||||||||          |  |             ||||||||||||||
Db    705 CATATGTTAACTGACCTAAGCGGCCATCACACCCTGCTTGCTTTGTGGTGGCTCATCA 764

Qy    605 CGGCCAGCCACCAAGACCATGAACTTTGGCTGCCAGAGAGCTGAGATCTTGGGAGCCCTG 664
          ||||||||          |  |             ||||||||||||||
Db    765 AAATCACCAACCAAAAAGATTCACCTTGGATTTCATCGCCTTAGAGGTTTTGTCAGCTATG 824

Qy    665 GTCTCTGTACTGTCCCATCTGGGTCGTGACGGGGGTACTGGTGGGTTACCTGGGGAGCGG 724
          ||||||||          |   |  ||||||||||||||||
Db    825 ATTAGTGTGCTGTTTGGGTGTATATACTTNATGGGATTCCCTTNTATPATGAAGCTGTGCAAAGA 884

Qy    725 CTGATCTCTGGGGGACTAGGAAANTTGACGGGGGGGACCATGCTGAATCACGTCGGGGCTGCGCT 784
          ||||||||          |        |   |||||||||||||
Db    885 ACTATCCATATGAACTATGAAATAATNGGGATANATAAATGCTCATCACCGCACAGCTGTTGGA 944

Qy    785 GTGGCTGTGAACATCATAAATNGGGGTTGACCCTTGACCCTTCACCAGTCTCTGGCCATGGGCACGCCAC 844
          ||||||||          |        |   |||||||||||||
Db    945 GTTGCAGTTAAATGTAATAAATNGGGGTTCTGTGTTGAAACCAGTCTCGGTCACGCGTCACTCCCAT 1004

Qy    845 GGCACCACC------------------AACCAGCAG 862
          ||||||||||||||||||||||||
Db   1005 TCCCACTCCCTGCCTTCANAATTCCCTACCACGAGGTTCTCGGGTGTGAACGTAACCATGGG 1064

Qy    863 GAGGAGAACCCCAGCGTCCGAGCTGCCCTTCACTTCATCCCATGTGATCGGGGCTTGATCCGGGAGC 922
          ||||||||          |        |   |||||||||||||
Db   1065 CAGGATAGCCTGGCCGAGTGAGAAGCTGCATTGTCATTTGTACATGCCTTGGGAGATCTGGTACGACAGT 1124

Qy    923 ATGGGTGTGCCTAGTGGCAGCCTATATTTATTTTTATACTTCAAGGCCAGAGAATACAAGTANGTACAC 982
          ||||||||          |        |   |||||||||||||
Db   1125 GTTGGGTGTGCTAATTAGCTGCCATNACATCACAAATACGANTTCAAGCCAGAATACAAGATTGCTGAC 1184
```

Qy   983 CCCATCTGCACCTTCGTCTCTCTCCATCCTGGTCTCCTGGGACAACCTTGACCCATCCTGAGA 1042
           ||||||||| |||| |||||||| || ||||||| | |||||||| |||||||||| ||| ||
Db   1185 CCCACTGTACATACGTATTTCATTCATTCATCTGTGGCTTTTACAACACATTTCGAATCATATGG 1244

Qy   1043 GATGTGATCCGTGTTGTTGATGGAAGGGACCCCCACAAGGGCGTTGACTTCACACCTGTTCGT 1102
           ||| |||||| | |||| || ||||| ||| || || |||| ||||| ||||||| ||
Db   1245 GATACAGTAGTTATAATACTAGAAGGTGTGCCAAGCCATTTGAATGTAGACTATATCAAA 1304

Qy   1103 GATCTCCTGCTGTCGGTGGGAGGGGTTGAGAAGCCCTGCCACAGCCTGCATATCTGGGCACTG 1162
           ||| |||||||||||||||| |||||| ||||||| |||||| |||||||| |||||||| ||
Db   1305 GAAGCCTTGATGAAAATAGAAGATGTTATATTCAGTCGGAAGATTTAAATATCTGGTCTCTC 1364

Qy   1163 ACGGTCGGCCCAGCCTGTTCTGTGTCCACAT 1194
           |||||||||||||| ||||| ||| ||||||
Db   1365 ACTTCAGGAAAATCTACTGCCATAGTTCACAT 1396

Query Match   4.6%; Score 73.8; DB 10; Length 278;
Best Local Similarity 59.8%; Pred. No. 9.4e-10;
Matches 140; Conservative   0; Mismatches   93; Indels   1; Gaps 1;

Qy   854 AACCACGAGGAGGAGAACCCCAGCGGTCCGAGCTGCCTTCATCCATGTGATCGGCGACTTT 913
           || || ||| ||||| ||| |||| |||||||| |||| ||||||||||||||||||| ||
Db   45 AAGAAGCCTCGGCGAACATCAACGTGCACACGCATACCTCCACGTGCTCGGGGACTCC 104

Qy   914 ATGCAGAGCCATGGGGTCGTGTCCTAGTGAGCAGCCTATATATTTATACTTCAAGCCAGAATACAAG 973
            ||||||||| ||||||||| || ||||||| || | |||||| || |||||| |||
Db   105 GTCCAGACGCGTCGGGGCGTCATGTGGCGGGGCCGGGCCGAGTGGAAG 164

Qy   974 TATGTAGACCCCATCTGCACCTTCTCTCTTCCATCTCTGGGGACAACCTTGACC 1033
            | ||| |||||||||||||||||||| |||||||||| ||||||||||| |||
Db   165 GTCATCGGACCTCATCCTGCCACCGTCGTCGTTCACGATCCGG 224

Qy   1034 ATCCTGAGAGATGTGATCCTCGTGTTGATGGAAGGGACCCCGGCGGTTGAC 1087
           |||||||||||| || ||||||| | |||| |||| |||||| |||| |||
Db   225 ATGCTCCG-CANATCCTCGAAGTGTTGATGAGAGCACGCCCGCGAGATCGAC 277

Query Match   3.8%; Score 61.8; DB 10; Length 900;
Best Local Similarity 71.7%; Pred. No. 2e-06;
Matches   81; Conservative   0; Mismatches   32; Indels   0; Gaps 0;

Qy   1501 TCCCAGCCTCCCATCTGACTACAGCCAGGGTGGGGACTCAGCGGGGTATAAAGCTAGTGTG 1560
           |||| |||| ||| || ||||||||| |||| || ||||| ||||||| |||||
Db   746 TCTCAGTCTCACCTGGAGCTACTGGGAGGGTAAAGCCATTTGAAGAATAAAGTCATCCAG 805

Qy   1561 ACCCTGAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 1613
           ||||||||||||||||||||||||||||||||||| ||||||||||||||||||||||||
Db   806 AGCCTCAAAAAAAAAAAAAAAAAAAATATAAAAAATAAAAAAAAAAAAAAAAAAAAAAAA 858

Query Match   3.8%; Score 61; DB 9; Length 1254;
Best Local Similarity 63.1%; Pred. No. 3.7e-06;
Matches   94; Conservative   0; Mismatches   55; Indels   0; Gaps 0;

Qy   1465 CTCCTGACCTCTGCCCACTCCAGGAATGGAGCTCTTCCCAGCCCTCCCATCTGACTACAG 1524
           ||| ||| || |||| ||| || |||| | ||| || ||| |||||||||||||||||||
Db   1049 CTCCAGATCTCAACCTGTTCCCTGGAAGTAGGGCCTGCTTCCATCCCAGTGAAATAAAC 1108

Qy   1525 CCAGGGTGGGGACTCACGCGGGGTATAAAGCTAGTGTGACCCTGACCCCAGTTATGGGA 1584
            || | ||||| ||| | |||||||| ||| |||||||| |||||||||||||||||||
Db   1109 ATGTATTAGACACCTACTGAGTATAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 1168

Qy   1585 AAAAAAAAAAAAAAAAAAAAAAAAAAAAA 1613
               |||||||||||||||||||||||||||||
Db   1169 AAAAAAAAAAAAAAAAAAAAAAAAAAAAA 1197

RESULT 8
US-10-033-350-1
; Sequence 1, Application US/10033350
; Publication No. US20030022284A1
; GENERAL INFORMATION:
;   APPLICANT: Genentech, Inc.
;   Klein, Robert D.
;   Moore, Mark W.
;   Rosenthal, Arnon
;   Ryan, Anne M.
;   NUMBER OF SEQUENCES: 12
;   CORRESPONDENCE ADDRESS:
;   ADDRESSEE: Genentech, Inc.
;   STREET: 1 DNA Way
;   CITY: South San Francisco
;   STATE: California
;   COUNTRY: USA
;   ZIP: 94080
;   COMPUTER READABLE FORM:
;   MEDIUM TYPE: 3.5 inch, 1.44 Mb floppy disk
;   COMPUTER: IBM PC compatible
;   OPERATING SYSTEM: PC-DOS/MS-DOS
;   SOFTWARE: WinPatin (Genentech)
;   CURRENT APPLICATION DATA:
;   APPLICATION NUMBER: US/10/033,350
;   FILING DATE: 02-No. US20030022284A1-2001
;   CLASSIFICATION: <Unknown>
;   PRIOR APPLICATION DATA:
;   APPLICATION NUMBER: US/08/860,370
;   FILING DATE: <Unknown>
;   APPLICATION NUMBER: PCT/US97/04363
;   FILING DATE: 13-Mar-1997
;   APPLICATION NUMBER: 08/615902
;   FILING DATE: 14-Mar-1996
;   APPLICATION NUMBER: 08/618236
;   FILING DATE: 14-Mar-1996
;   ATTORNEY/AGENT INFORMATION:
;   NAME: Torchia, PhD., Timothy E.
;   REGISTRATION NUMBER: 36,700
;   REFERENCE/DOCKET NUMBER: P0996P1PCT
;   TELECOMMUNICATION INFORMATION:
;   TELEPHONE: 650/225-8674
;   TELEFAX: 650/952-9881
;   INFORMATION FOR SEQ ID NO: 1:
;   SEQUENCE CHARACTERISTICS:
;   LENGTH: 2378 base pairs
;   TYPE: Nucleic Acid
;   STRANDEDNESS: Double
;   TOPOLOGY: Linear
;   SEQUENCE DESCRIPTION: SEQ ID NO: 1:
US-10-033-350-1

Query Match   3.6%; Score 58.6; DB 9; Length 2378;
Best Local Similarity   62.8%; Pred. No. 2e-05; Matches 54; Indels 0; Gaps 0;
Matches 91; Conservative 0; Mismatches 54; Indels 0; Gaps 0;

Qy   1469 TGACCTCTGCCCACTCCAGGAATGGAGCTCTTCCCAGCCTCCCATCTGACTACAGCCAG 1528
               ||| ||| | ||||| ||| |||||||| ||| |||||||||||||||||||||||||
Db   2156 TGGCCTGGGCACAGCTCCCGGATTGTTGATGGGCACTCTCCTGAGCTTTGCTTGAGTGAG 2215

Qy   1529 GGTGGGGACTCAGCGCGGGTATAAAGCTAGTGTGACCCTGAAAAAAAAAAAAAAAAAAAA 1588
               ||| || | | |||| ||||||||||||||||||||||||||||||||||||||||||||
Db   2216 AAGCTGAATGTAGCTGAAAATCAACTCTTCTTACACTTAAAAAAAAAAAAAAAAAAAAAA 2275

Qy   1589 AAAAAAAAAAAAAAAAAAAAAAAA 1613
               ||||||||||||||||||||||||
Db   2276 AAAAAAAAAAAAAAAAAAAAAAAA 2300

RESULT 9
US-10-091-483-54
; Sequence 54, Application US/10091483
; Publication No. US20030049650A1
; GENERAL INFORMATION:
;   APPLICANT: Rosen et al.
;   TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
;   FILE REFERENCE: PT212C1
;   CURRENT APPLICATION NUMBER: US/10/091,483
;   CURRENT FILING DATE: 2002-03-07
;   NUMBER OF SEQ ID NOS: 348
;   SOFTWARE: Patentin Ver. 2.0
;   Prior Application removed - See File Wrapper or Palm
; SEQ ID NO 54
;   LENGTH: 307
;   TYPE: DNA
;   ORGANISM: Homo sapiens
US-10-091-483-54

Query Match   3.6%; Score 58.2; DB 9; Length 307;
Best Local Similarity   88.7%; Pred. No. 1.1e-05; Matches 63; Indels 0; Gaps 0;
Matches 63; Conservative 0; Mismatches 8; Indels 0; Gaps 0;

Qy   1543 GGGTATAAAGCTAGTGTGACCCTGAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 1602
               | |||||||||| | ||| | || ||||||||||||||||||||||||||||||||||||
Db   186 GTGAATAAAGCTGCTCTGAGCATTAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 245

Qy   1603 AAAAAAAAAAA 1613
               |||||||||||
Db   246 AAAAAAAAAAA 256

RESULT 10
US-09-764-846-54
; Sequence 54, Application US/09764846
; Patent No. US20020102638A1
; GENERAL INFORMATION:
;   APPLICANT: Rosen et al.
;   TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
;   FILE REFERENCE: PT212
;   CURRENT APPLICATION NUMBER: US/09/764,846
;   CURRENT FILING DATE: 2001-01-17
;   Prior application data removed - consult PALM or file wrapper
;   NUMBER OF SEQ ID NOS: 348
;   SOFTWARE: Patentin Ver. 2.0
; SEQ ID NO 54
;   LENGTH: 307
;   TYPE: DNA
;   ORGANISM: Homo sapiens
US-09-764-846-54

Query Match   3.6%; Score 58.2; DB 10; Length 307;
Best Local Similarity   88.7%; Pred. No. 1.1e-05; Matches 63; Indels 0; Gaps 0;
Matches 63; Conservative 0; Mismatches 8; Indels 0; Gaps 0;

Qy   1543 GGGTATAAAGCTAGTGTGACCCTGAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 1602
               | |||||||||| | ||| | || ||||||||||||||||||||||||||||||||||||
Db   186 GTGAATAAAGCTGCTCTGAGCATTAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 245

Qy   1603 AAAAAAAAAAA 1613
               |||||||||||
Db   246 AAAAAAAAAAA 256

RESULT 11
US-10-091-483-126
; Sequence 126, Application US/10091483
; Publication No. US20030049650A1
; GENERAL INFORMATION:
;   APPLICANT: Rosen et al.
;   TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies

```
          FILE REFERENCE: PT212C1
          CURRENT APPLICATION NUMBER: US/10/091,483
          CURRENT FILING DATE: 2002-03-07
          NUMBER OF SEQ ID NOS: 348
          Prior Application removed - See File Wrapper or Palm
          SOFTWARE: Patentin Ver. 2.0
          SEQ ID NO 126
          TYPE: DNA
          LENGTH: 314
          ORGANISM: Homo sapiens
          FEATURE:
          NAME/KEY: misc_feature
          LOCATION: (303)
          OTHER INFORMATION: n equals a,t,g, or c
          NAME/KEY: misc_feature
          LOCATION: (309)
          OTHER INFORMATION: n equals a,t,g, or c
          NAME/KEY: misc_feature
          LOCATION: (314)
          OTHER INFORMATION: n equals a,t,g, or c
          US-10-091-483-126

Qy 1543 GGGTATAAAGCTAGTGTGTGCACCCTGAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 1602
        |-||||||||  ||||  |  |  ||||||||||||||||||||||||||||||||||||||||
Db  181 GTGAATAAAGCTGCTCTCGAGCATTAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 240

Qy 1603 AAAAAAAAAA 1613
        ||||||||||
Db  241 AAAAAAAAAA 251


RESULT 12
US-09-764-846-126
Sequence 126, Application US/09764846
Patent No. US20020102638A1
GENERAL INFORMATION:
APPLICANT: Rosen et al.
TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
FILE REFERENCE: PT212
CURRENT APPLICATION NUMBER: US/09/764,846
CURRENT FILING DATE: 2001-01-17
Prior application data removed - consult PALM or file wrapper
NUMBER OF SEQ ID NOS: 348
SOFTWARE: Patentin Ver. 2.0
SEQ ID NO 126
LENGTH: 314
TYPE: DNA
ORGANISM: Homo sapiens
FEATURE:
NAME/KEY: SITE
LOCATION: (303)
OTHER INFORMATION: n equals a,t,g, or c
NAME/KEY: SITE
LOCATION: (309)
OTHER INFORMATION: n equals a,t,g, or c
NAME/KEY: SITE
LOCATION: (314)
OTHER INFORMATION: n equals a,t,g, or c
US-09-764-846-126

Query Match    3.6%;   Score 58.2;   DB 9;   Length 314;
Best Local Similarity   88.7%;   Pred. No. 1.1e-05;
Matches  63;   Conservative  0;   Mismatches  8;   Indels  0;   Gaps  0;

Qy 1543 GGGTATAAAGCTAGTGTGTGCACCTGAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 1602
        |-||||||||| ||||  |  | ||||||||||||||||||||||||||||||||||||||||
Db  181 GTGAATAAAGCTGCTCTGAGCATTAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 240

Qy 1603 AAAAAAAAAA 1613
        ||||||||||


Query Match    3.6%;   Score 58.2;   DB 10;   Length 314;
Best Local Similarity   88.7%;   Pred. No. 1.1e-05;
Matches  63;   Conservative  0;   Mismatches  8;   Indels  0;   Gaps  0;

Qy 1543 GGGTATAAAGCTAGTGTGTGCACCTGAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 1602
        |-||||||||| ||||  |  | ||||||||||||||||||||||||||||||||||||||||
Db  181 GTGAATAAAGCTGCTCTGAGCATTAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 240

Qy 1603 AAAAAAAAAA 1613
```

```
Db  241 AAAAAAAAAA 251


RESULT 13
US-09-925-299-16
Sequence 16, Application US/09925299
Publication No. US20030040617A9
GENERAL INFORMATION:
APPLICANT: Rosen et al.
TITLE OF INVENTION: Nucleic Acids, Proteins and Antibodies
FILE REFERENCE: PA102
CURRENT APPLICATION NUMBER: US/09/925,299
CURRENT FILING DATE: 2001-08-10
PRIOR APPLICATION NUMBER: PCT/US00/05883
PRIOR FILING DATE: 2000-03-08
PRIOR APPLICATION NUMBER: 60/124,270
PRIOR FILING DATE: 1999-03-12
NUMBER OF SEQ ID NOS: 1556
SOFTWARE: Patentin Ver. 2.0
SEQ ID NO 16
LENGTH: 371
TYPE: DNA
ORGANISM: Homo sapiens
FEATURE:
NAME/KEY: misc_feature
LOCATION: (350)
OTHER INFORMATION: n equals a,t,g, or c
NAME/KEY: misc_feature
LOCATION: (360)
OTHER INFORMATION: n equals a,t,g, or c
US-09-925-299-16

Query Match    3.5%;   Score 57.4;   DB 9;   Length 371;
Best Local Similarity   91.0%;   Pred. No. 1.9e-05;
Matches  61;   Conservative  0;   Mismatches  6;   Indels  0;   Gaps  0;

Qy 1547 ATAAAGCTAGTGTGTGCACCCTGAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 1606
        ||||||||| | ||||||  |  | |||||||||||||||||||||||||||||||||||||||
Db  251 ATAAAGCTCTACTGTCTCGAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 310

Qy 1607 AAAAAAAA 1613
        ||||||||
Db  311 AAAAAAAA 317


RESULT 14
US-09-925-299-16
Sequence 16, Application US/09925299
Patent No. US20020055627A1
GENERAL INFORMATION:
APPLICANT: Rosen et al.
TITLE OF INVENTION: Nucleic Acids, Proteins and Antibodies
FILE REFERENCE: PA102
CURRENT APPLICATION NUMBER: US/09/925,299
CURRENT FILING DATE: 2001-08-10
PRIOR APPLICATION NUMBER: PCT/US00/05883
PRIOR FILING DATE: 2000-03-08
PRIOR APPLICATION NUMBER: 60/124,270
PRIOR FILING DATE: 1999-03-12
NUMBER OF SEQ ID NOS: 1556
SOFTWARE: Patentin Ver. 2.0
SEQ ID NO 16
LENGTH: 371
TYPE: DNA
ORGANISM: Homo sapiens
FEATURE:
NAME/KEY: misc_feature
LOCATION: (350)
OTHER INFORMATION: n equals a,t,g, or c
NAME/KEY: misc_feature
LOCATION: (360)
OTHER INFORMATION: n equals a,t,g, or c
```

US-09-925-299-16

```
Qy   Query Match          3.5%;  Score 57.4;  DB 10;  Length 371;
     Best Local Similarity  91.0%;  Pred. No. 1.9e-05;
     Matches  61; Conservative   0; Mismatches   6; Indels   0; Gaps   0;

Qy   1547 ATAAAGCTAGTGTGACCCTGAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 1606
          ||||||||   | | ||||||||||||||||||||||||||||||||||||||||||||
Db    251 ATAAAGCTCTACTGTCTCTGAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 310

Qy   1607 AAAAAAA 1613
          |||||||
Db    311 AAAAAAA 317
```

RESULT 15

US-09-745-763-149

Sequence 149, Application US/09745763

Patent No. US2002006594A1

GENERAL INFORMATION:

~   APPLICANT: Jacobs, Kenneth
~   McCoy, John M.
~   LaValie, Edward R.
~   Collins-Racie, Lisa A.
~   Evans, Cheryl
~   Merberg, David
~   Treacy, Maurice
~   Spaulding, Vikki

~   TITLE OF INVENTION: SECRETED PROTEINS AND POLYNUCLEOTIDES
~   ENCODING THEM

~   NUMBER OF SEQUENCES: 219

~   CORRESPONDENCE ADDRESS:
~   ADDRESSEE: Genetics Institute, Inc.
~   STREET: 87 CambridgePark Drive
~   CITY: Cambridge
~   STATE: MA
~   COUNTRY: U.S.A.
~   ZIP: 02140

~   COMPUTER READABLE FORM:
~   MEDIUM TYPE: Floppy disk
~   COMPUTER: IBM PC compatible
~   OPERATING SYSTEM: PC-DOS/MS-DOS
~   SOFTWARE: Patentin Release #1.0, Version #1.30

~   CURRENT APPLICATION DATA:
~   APPLICATION NUMBER: US/09/745,763
~   FILING DATE: 18-Jun-2000
~   CLASSIFICATION: <Unknown>

~   ATTORNEY/AGENT INFORMATION:
~   NAME: Sprunger, Suzanne A.
~   REGISTRATION NUMBER: 41,323

~   TELECOMMUNICATION INFORMATION:
~   TELEPHONE: (617) 498-8284
~   TELEFAX: (617) 876-5851

~   INFORMATION FOR SEQ ID NO: 149:
~   SEQUENCE CHARACTERISTICS:
~   LENGTH: 1152 base pairs
~   TYPE: nucleic acid
~   STRANDEDNESS: double
~   TOPOLOGY: linear
~   MOLECULE TYPE: cDNA
~   SEQUENCE DESCRIPTION: SEQ ID NO: 149:

US-09-745-763-149

```
Qy   Query Match          3.5%;  Score 57;  DB 10;  Length 1152;
     Best Local Similarity  77.5%;  Pred. No. 3.9e-05;
     Matches  69; Conservative   0; Mismatches  20; Indels   0; Gaps   0;

Qy   1525 CCAGGGTGGGGACTCAGCGGGTATAAAGCTAGTGTGACCCTGAAAAAAAAAAAAAAAA 1584
          | | | | ||| | ||    | ||  ||||| ||||||||||||||||||||||||||
Db   1028 CGAGAGCGCGGGGCTCAAGGGCAATAAAGCAGCTCCACTCTCAAAAAAAAAAAAAAAA 1087

Qy   1585 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 1613
          ||||||||||||||||||||||||||||||
```

Db   1088 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 1116

Search completed: March 30, 2003, 01:54:56
Job time : 191.281 secs

OM nucleic - nucleic search, using sw model

GenCore version 5.1.4_p5_4578
Copyright (c) 1993 - 2003 Compugen Ltd.

Run on: March 29, 2003, 13:11:19 ; Search time 3108.9 Seconds
(without alignments)
9361.127 Million cell updates/sec

Title: US-09-691-219-3_COPY_1_1000
Perfect score: 1000
Sequence: 1 cctgccacatgcctggcta.........caagaggcctgcacctctga 1000
Scoring table: IDENTITY_NUC
                Gapop 10.0 , Gapext 1.0
Searched: 2054640 seqs, 1455140287B residues
Total number of hits satisfying chosen parameters: 4109280
Post-processing: Minimum Match 0%
                 Maximum Match 100%
Minimum DB seq length: 0
Maximum DB seq length: 2000000000
                 Listing first 45 summaries

Database : GenEmbl:*
         1: gb_ba:*
         2: gb_htg:*
         3: gb_in:*
         4: gb_om:*
         5: gb_ov:*
         6: gb_pat:*
         7: gb_ph:*
         8: gb_pl:*
         9: gb_pr:*
        10: gb_ro:*
        11: gb_sts:*
        12: gb_sy:*
        13: gb_un:*
        14: gb_vi:*
        15: em_ba:*
        16: em_fun:*
        17: em_hum:*
        18: em_in:*
        19: em_mu:*
        20: em_om:*
        21: em_or:*
        22: em_ov:*
        23: em_pat:*
        24: em_ph:*
        25: em_pl:*
        26: em_ro:*
        27: em_sy:*
        28: em_un:*
        29: em_vi:*
        30: em_htg_hum:*
        31: em_htg_inv:*
        32: em_htg_other:*
        33: em_htg_mus:*
        34: em_htg_pln:*
        35: em_htg_rod:*
        36: em_htg_mam:*
        37: em_htg_vrt:*
        38: em_sy:*
        39: em_htgo_hum:*
        40: em_htgo_mus:*
        41: em_htgo_other:*

Pred. No. is the number of results predicted by chance to have a

score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
	1	1000	100.0	11101	6	AX411375	AX411375 Sequence
	2	1000	100.0	176006	9	AL391650	AL391650 Human DNA
C	3	244.4	24.4	194191	2	AC097332	AC097332 Pan trogl
C	4	234.6	23.5	32289	9	AC093822	AC093822 Homo sapi
	5	230	23.0	186625	9	AC006487	AC006487 Homo sapi
	6	228	22.8	170862	9	AC015795	AC015795 Homo sapi
	7	227.6	22.8	143335	9	AC025188	AC025188 Homo sapi
C	8	226.6	22.7	110965	2	AC020927	AC020927 Homo sapi
C	9	226.6	22.7	178451	9	AL139396	AL139396 Human DNA
	10	226.4	22.6	167924	9	AC067891	AC067891 Homo sapi
	11	226	22.6	129043	9	AL132712	AL132712 Human chr
	12	225.6	22.6	37784	9	AC002997	AC002997 Human DNA
C	13	225.6	22.6	98713	9	AC025511	AC025511 Human DNA
	14	225.6	22.6	172203	9	AC005863	AC005863 Homo sapi
	15	225.2	22.5	44183	2	AC093229	AC093229 Homo sapi
	16	225.2	22.5	110233	9	AC124853	AC124853 Homo sapi
	17	224.4	22.4	142728	9	AC005858	AC005858 Homo sapi
	18	224.4	22.4	149913	2	AC010330	AC010330 Homo sapi
	19	223.6	22.4	157463	9	AC091529	AC091529 Homo sapi
	20	223.6	22.4	286758	9	AC006449	AC006449 Homo sapi
	21	223.4	22.3	187984	2	AP001262	AP001262 Homo sapi
	22	223	22.3	16781	9	AL356422	AL356422 Human DNA
C	23	222.4	22.2	150336	9	H53395C13	H53395C13 Human DNA
C	24	222.2	22.2	88665	2	AL162661	AL162661 Human DNA
C	25	222.2	22.2	152121	2	AP000830	AP000830 Homo sapi
	26	222.2	22.2	194310	9	AP001267	AP001267 Homo sapi
	27	221.8	22.2	146017	2	AC027473	AC027473 Homo sapi
	28	221.2	22.1	39673	9	AL672238	AL672238 Human DNA
	29	221.2	22.1	87375	9	AP000289	AP000289 Homo sapi
	30	221.2	22.1	100000	2	AP000042	AP000042 Homo sapi
	31	221.2	22.1	100000	9	AP000110	AP000110 Homo sapi
	32	221.2	22.1	100000	9	AP000186	AP000186 Homo sapi
C	33	221.2	22.1	209856	9	AC090543	AC090543 Homo sapi
C	34	221.2	22.1	222759	2	AC073177	AC073177 Homo sapi
	35	221.2	22.1	340000	9	AP001716	AP001716 Homo sapi
C	36	220.8	22.1	25233	9	AL591212	AL591212 Human DNA
C	37	220.6	22.1	145050	9	AC021078	AC021078 Homo sapi
	38	220.4	22.0	144555	2	AC037451	AC037451 Homo sapi
C	39	220.4	22.0	189536	9	AC117834	AC117834 Human DNA
	40	220	22.0	191722	2	AC116360	AC116360 Homo sapi
	41	220	22.0	168006	2	AC126227	AC126227 Papio cyn
C	42	219.8	21.9	181386	9	AC076686	AC076686 Homo sapi
C	43	219.2	21.9	150804	9	AC018500	AC018500 Homo sapi
	44	219	21.9	150804	9	AC018500	AC018500 Homo sapi
	45	218.8	21.9	187127	9	AC097359	AC097359 Homo sapi

RESULT 1
AX411375

LOCUS   AX411375   11101 bp   DNA   linear   PAT 14-JUN-2002
DEFINITION Sequence 3 from Patent WO0224910.
ACCESSION   AX411375
VERSION   AX411375.1 GI:21444025
KEYWORDS   .
SOURCE   human.
  ORGANISM Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1
  AUTHORS   Wei,M.H., Ketchum,K.A., di Francesco,V. and Beasley,E.M.
  TITLE   Isolated human transporter proteins, nucleic acid molecules
            encoding them, and uses thereof

ALIGNMENTS

JOURNAL   Patent: WO 0224910-A 3 28-MAR-2002;
          PE Corporation (NY) (US)
FEATURES   Location/Qualifiers
   source   1..11101
                /organism="Homo sapiens"
                /db_xref="taxon:9606"
ORIGIN

BASE COUNT   2403 a   3041 c   3200 g   2448 t   9 others

Query Match
Best Local Similarity   100.0%;   Score 1000;   DB 6;   Length 11101;
Matches 1000;   Conservative   0;   Pred. No. 1.8e-269;   Mismatches   0;   Indels   0;   Gaps   0;

```
Qy       1 CCTGCCACCATCCTGCCTAATTTTTCTTATTTTTTAGTTAGRAGACGAGGGTTTTTGCCATGTTG    60
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1 CCTGCCACCATCCTGCCTAATTTTTCTTATTTTTTAGTTAGRAGACGAGGGTTTTTGCCATGTTG    60

Qy      61 ACCAGCCTGGTCTCGAACTCTTGACCTCAGGTGATCCGCCTGCCTCAGCCTCCCAAAGTG   120
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      61 ACCAGCCTGGTCTCGAACTCTTGACCTCAGGTGATCCGCCTGCCTCAGCCTCCCAAAGTG   120

Qy     121 CTGGGATTATAGGCGTGAGCCGCCGCGCCCACCAGCCCACCAGCCCACCAAACTGAAAAGTAG   180
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     121 CTGGGATTATAGGCGTGAGCCGCCGCGCCCACCAGCCCACCAGCCCACCAAACTGAAAAGTAG   180

Qy     181 AGGGAATAAGTTATAGTGTATCACCCCATTTACCCCATCACTCAGTTTCAACAGCTGTTGACATA   240
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     181 AGGGAATAAGTTATAGTGTATCACCCCATTTACCCCATCACTCAGTTTCAACAGCTGTTGACATA   240

Qy     241 TTTATTTCTTCTTCTATACCAGTACCGTACCTACTCTCCCCCACTGGGATTAATTTTAAGGCAAAACCC   300
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     241 TTTATTTCTTCTTCTATACCAGTACCGTACCTACTCTCCCCCACTGGGATTAATTTTAAGGCAAAACCC   300

Qy     301 AGATGACATTTTATCCCTAAATACTTTTAGATANAAGGTGTTCTTTGAAAAAAATCATAACC   360
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     301 AGATGACATTTTATCCCTAAATACTTTTAGATANAAGGTGTTCTTTGAAAAAAATCATAACC   360

Qy     361 TCAGGACCAGCCTGGCCAACANGGTGAAAACCCTGTCTCGTGACTTTACTAAAAATACAAAAAATTAGC   420
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     361 TCAGGACCAGCCTGGCCAACANGGTGAAAACCCTGTCTCGTGACTTTACTAAAAATACAAAAAATTAGC   420

Qy     421 TTGGCATGGTGGCGTGGGGCACCTGTAATCCCAGCTACTCAGGAGGCTGAGGCAGGAGAATCA   480
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     421 TTGGCATGGTGGCGTGGGGCACCTGTAATCCCAGCTACTCAGGAGGCTGAGGCAGGAGAATCA   480

Qy     481 CTTGAATCCGGGAAGCAGAGATTGCAGTGAGCTGAGATTGCAGCCTGGGCGACAG   540
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     481 CTTGAATCCGGGAAGCAGAGATTGCAGTGAGCTGAGATTGCAGCCTGGGCGACAG   540

Qy     541 AGACAGAAATGAAAACTCTGTCTCAAAAAAACAAACAAACAACTATACATATAAAA   600
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     541 AGACAGAAATGAAAACTCTGTCTCAAAAAAACAAACAAACAACTATACATATAAAA   600

Qy     601 TGAACAAATGANTGCCACAATAGCACCAGAGAATTTTATAAATACAGATTCCCAGCCCCTGC   660
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     601 TGAACAAATGANTGCCACAATAGCACCAGAGAATTTTATAAATACAGATTCCCAGCCCCTGC   660

Qy     661 CCCAGACCTACTGAATCCTGGAAATATTCAGGCTCCACACCCAGAGATTCTGGTTCGGTT   720
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     661 CCCAGACCTACTGAATCCTGGAAATATTCAGGCTCCACACCCAGAGATTCTGGTTCGGTT   720

Qy     721 GGTCGATGCAGGGACCTGTAACCTGCGTTGTAACACTTCCAGGTAAATCCTGACCCT   780
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     721 GGTCTGATGCAGGGACCTGTAACCTGCGTTGTAACACTTCCAGGTAAATCCTGACCCT   780

Qy     781 GCTGGTGCTTCAGAGTRAGACANCAGAGACCTGGAGAAAACCAGAGGTGTGTTCAGAGA   840
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     781 GCTGGTGCTTCAGAGTAGACACAGAGACCTGGAGAAAACCAGAGGTGTGTTCAGAGA   840

Qy     841 AAACCAGAGTCCAGAGAAGCAGCACTCAGTGAGGACCCAAGCAGAGAGGGGGTGCAC   900
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     841 AAACCAGAGTCCAGAGAAGCAGCACTCAGTGAGGACCCAAGCAGAGAGGGGGTGCAC   900
```

RESULT 2
AL391650/c
LOCUS   AL391650   176006 bp   DNA   linear   PRI   28-NOV-2001
DEFINITION   Human DNA sequence from clone RP11-96L14 on chromosome 1, complete
           sequence.
ACCESSION   AL391650
VERSION   AL391650.18   GI:16944045
KEYWORDS   HTG.
SOURCE   human.
  ORGANISM   Homo sapiens
           Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
           Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1 (bases 1 to 176006)
  AUTHORS   Heath.P.
  TITLE   Direct Submission
  JOURNAL   Submitted (28-NOV-2001) Wellcome Trust Sanger Institute, Hinxton,
           Cambridgeshire, CB10 1SA, UK. E-mail enquiries: clonerequest@sanger.ac.uk
           humquery@sanger.ac.uk Clone requests: clonerequest@sanger.ac.uk
COMMENT   On Nov 15, 2001 this sequence version replaced gi:16605685.
           During sequence assembly data is compared from overlapping clones.
           Where differences are found these are annotated as variations
           together with a note of the overlapping clone name. Note that the
           variation annotation may not be found in the sequence submission
           corresponding to the overlapping clone, as we submit sequences with
           only a small overlap as described above.
           This sequence was finished as follows unless otherwise noted: all
           regions were either double-stranded or sequenced with an alternate
           chemistry or covered by high quality data (i.e., phred quality >=
           30); an attempt was made to resolve all sequencing problems, such
           as compressions and repeats; all regions were covered by at least
           one plasmid subclone or more than one M13 subclone; and the
           assembly was confirmed by restriction digest. The following
           abbreviations are used to associate primary accession numbers given
           in the feature table with their source databases: Em:, EMBL; Sw:,
           SWISSPROT; Tr:, TREMBL; Wp:, WORMPEP; Information on the WORMPEP
           database can be found at
           http://www.sanger.ac.uk/Projects/C_elegans/wormpep This sequence
           was generated from part of bacterial clone contigs of human
           chromosome 1, constructed by the Sanger Centre Chromosome 1 Mapping
           Group. Further information can be found at
           http://www.sanger.ac.uk/HGP/Chr1
           RP11-96L14 is from the library RPCI-11.1 constructed by the group
           of Pieter de Jong. For further details see
           http://www.chori.org/bacpac/home.htm
           VECTOR: pBACe3.6
           This sequence is the entire insert of clone RP11-96L14 The true
           left end of clone RP11-231P20 is at 72881 in this sequence. The
           true right end of clone RP11-111D20 is at 72886 in this sequence.
FEATURES   Location/Qualifiers
   source   1..176006
                /organism="Homo sapiens"
                /db_xref="taxon:9606"
                /chromosome="1"
                /clone="RP11-96L14"
                /clone_lib="RPCI-11.1"
   misc_feature   169390..169425
                /note="sequence from uni-directional dGTP big dye
                terminator reads only."
ORIGIN

BASE COUNT   45546 a   43652 c   43353 g   43455 t

Query Match
Best Local Similarity   100.0%;   Score 1000;   DB 9;   Length 176006;
Matches 1000;   Conservative   0;   Pred. No. 2.5e-269;   Mismatches   0;   Indels   0;   Gaps   0;

```
Qy     901 CTCACATCCTCACTCCTGCGCACCCGTCTCCCTACAAGATGATGGAGGACTGAGAGAGCCCCTT   960
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     901 CTCACATCCTCACTCCTGCCACCCGTCTCCCTACAAGATGAGAGGACTGAGAGAGCCCCTT   960

Qy     961 CCTGTCCCCAGTGGTGGTGTGGGCAAGAGGCCTGCACCTCTCGA   1000
           ||||||||||||||||||||||||||||||||||||||||||||
Db     961 CCTGTCCCCAGTGGTGGTGTGGGCAAGAGGCCTGCACCTCTGA   1000
```

```
Qy        1 CCTGCCACCATGCTGCTTCGGCTAATTTTCTTATTTTTAGTAGAGACGAGGTTTGCCATGTTG      60
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    33679 CCTGCCACCATGCTGCTTCGGCTAATTTTCTTATTTTTAGTAGAGACGAGGTTTGCCATGTTG   33620

Qy       61 ACCAGGCTGGTCTCGAACTCTTGACCTCAGGTGATCCGCCCGCCTCAGCCTCCCAAAGTG     120
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    33619 ACCAGGCTGGTCTCGAACTCTTGACCTCAGGTGATCCGCCCGCCTCAGCCTCCCAAAGTG   33560

Qy      121 CTGGGATTATAGGCGTGAGCCGCCGCCACCCAGCCACACATTTTTTAAATACTGAAAAGTAG     180
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    33559 CTGGGATTATAGGCGTGAGCCGCCGCCACCCAGCCACACATTTTTTAAATACTGAAAAGTAG   33500

Qy      181 AGGGAATAGTAGTATAGTGTACCCCATTTACCCATTCACACATTTTTTAAATACTGAAAGTAG     240
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    33499 AGGGAATAGTAGTATAGTGTACCCCATTTACCCATTCACACATTTTTTCAACAGCTGGTGACATA   33440

Qy      241 TTTTATTTCTTCTTCTATACCAGTACCGTACTCTCCCCATTCACACATTTTTTAAGGCAAAACC     300
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    33499 TTTTATTTCTTCTTCTATACCAGTACCGTACTCTCCCCATCACACATTTTTTAAGGCAAAACC   33440

Qy      301 AGATGACATTTTATCCCTTAAAATACTTTAGATAAAGGTGTTCTTTTGAAAAATCATAACC     360
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    33439 AGATGACATTTTATCCCTTAAAATACTTTAGATAAAGGTGTTCTTTTGAAAAATCATAACC   33380

Qy      361 TCAGGACCAGCCTGGCCAACATGGTGAAACCCTGTCTCTACTAAAAATACAAAAATTAGC     420
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    33379 TCAGGACCAGCCTGGCCAACATGGTGAAACCCTGTCTCTACTAAAAATACAAAAATTAGC   33320

Qy      421 TTGGGCCATGGTCGTGTGGGCACCTGTAATCCCAGCTACTCAGGAGGCAGGAGAATCA     480
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    33319 TTGGGCCATGGTCGTGTGGGCACCTGTAATCCCAGCTACTCAGGAGGCAGGAGAATCA   33260

Qy      481 CTTGAACCCGGGAAGCAGAGAGAGTTGCAGTGAGCTGAGATTGCAGTGAGCCTGGGCGACAG     540
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    33259 TTGGCAATGGTCGTGGGCGACCTGTAATCCCAGCTACTCGGGAGGCTGAGGCAGGAGAATCA   33200

Qy      541 AGACAGAGAAATGAAACTCTGTCTCAAAAAACAAACAAACAAACAACTATACATAAAAA     600
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    33199 CTTGAACCCGGGAAGCAGAGAGATTGCAGTGAGCTGAGATCGCGACCTGCACTCCAGCCTGC   33140

Qy      601 TGAACAATGATGCCACAATAGCACCAGAATTTATAAATACAGAGATTCCCAGCCCTGC     660
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    33139 AGACAGAGAAATGAAACTCTGTCTCAAAAAACAAACAAACAAACAACTATACATAAAAA   33080

Qy      661 CCCAGACCTACTGAATCCTGGAAAATATTCAGCGTCCCAGAGAGATTCTGGTTCGGTT     720
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    33079 TGAACAATGATGCCACAATAGCACCAGAATTTATAAATACAGATTCCCAGCCCTGC   33020

Qy      721 GGTCTGAGTGCAGGGACCTGTAACCTGGGTGTTGTAACACCTTTCTCCAGGTAATGCTGAGCCT     780
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    33019 CCCAGACCTACTGAATCCTGGAAATATTCAGCGTCCCAGAGAGATTCTGGTTCGGTT   32960

Qy      781 GCTGGGTGGCTTCAGAGTAGAGACAGACACCTGGAGAAAAAACCAGGGTGTGTCTGAGGTTTCCAGAAGA     840
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    32959 GGTCTGAGTGCAGGGACCTGTAACCTGGGTGTTGTAACACCTTTCTCCAGGTAATGCTGAGCCT   32900

Qy      841 AAACCAGAGTCCAGAGAAGAGCAGAGAGGCACTGGAGGACCCAAGAGAGAGCGGGTGCAC     900
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    32899 GCTGGGTGGCTTCAGAGTAGAGACAGACACCTGGAGAAAAAACCAGGGTGTGTCTGAGGTTTTCCAGAAGA   32840

Qy      901 CTCACAATCCTCACTCCTGGCACCCCGTCTCCTACAAGATGAGAGACTGAAAGAGCCCCTT     960
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    32839 AAACCAGAGTCCAGAGAAGAGCAGAGAGGCACTGGAGGACCCAAGAGAGAGCGGGTGCAC   32780

Qy      961 CCTGTCCCCAGTGGTGGTGTGGGCAAGAGGGCCTGCACCTCTGA    1000
            ||||||||||||||||||||||||||||||||||||||||||||
Db    32779 CTCACAATCCTCACTCCTGGCACCCCGTCTCCTACAAGATGAGAGACTGAAAGAGCCCCTT   32720

Db    32719 CCTGTCCCCAGTGGTGGTGTGGGCAAGAGGGCCTGCACCTCTGA   32680
```

```
                        Summary Statistics
          Sequencing vector: Plasmid; M77789
          Chemistry: Dye-terminator Big Dye; 100% of reads
          Assembly program: Phrap; version 0.990329
          Consensus quality: 193448 bases at least Q40
          Consensus quality: 193860 bases at least Q30
          Consensus quality: 194094 bases at least Q20
          Estimated insert size: 194270; sum-of-contigs estimation
          Quality coverage: 0x in Q20 bases; agarose-fp estimation
          Quality coverage: 8.5x in Q20 bases; sum-of-contigs estimation
  --------------
  * NOTE: Estimated insert size may differ from sequence length
  *   (see http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html).
  * NOTE: This is a 'working draft' sequence. It currently
  *   consists of 5 contigs. The true order of the pieces
  *   is not known and their order in this sequence record is
  *   arbitrary. Gaps between the contigs are represented as
  *   runs of N, but the exact sizes of the gaps are unknown.
  *   this record will be updated with the finished sequence
  *   as soon as it is available and the accession number will
  *   be preserved.
  *        1      75617: contig of 75617 bp in length
  *    75618      75717: gap of unknown length
  *    75718     132426: contig of 56709 bp in length
  *   132427     132526: gap of unknown length
  *   132527     180942: contig of 48416 bp in length
  *   180943     181042: gap of unknown length
  *   181043     183092: contig of 2050 bp in length
  *   183093     183192: gap of unknown length
  *   183193     194191: contig of 10999 bp in length

FEATURES             Location/Qualifiers
     source          1..194191
                     /organism="Pan troglodytes"
                     /db_xref="taxon:9598"
                     /clone="RP43-22B16"
BASE COUNT    53320 a   44231 c   44666 g   51574 t   400 others
ORIGIN

Query Match      24.4%;  Score 244.4;  DB 2;  Length 194191;
Best Local Similarity 57.8%;  Pred. No. 2e-57;
Matches 349; Conservative 0; Mismatches 251; Indels 4; Gaps 1;

Qy         1 CCTGCCACCATGCCTGGCTAATTTTCTTATTTTAGTAGAGACGAGGTTTTGCCATGTTG     60
             ||                |||| ||||||| || |||||||||||||  ||||| ||||||
Db    132731 CCCACCACCATGCCTGGCTAATTTTTTGTATTTTAGTAGAGACGGGGTTTCACCATGTTG 132672

Qy        61 ACCAGGCTGGTCTCGAACTCCTGACCTCAGGTGATCCGCCGCCTCGCCTCCCAAAGTG    120
             ||||||||||||||||||||||||||||||||||| ||||  ||||||||||||||| |
Db    132671 GCCAGGCTGGTCTCGAACTGGTCAGCCGCCCACCAGCC--AACATTTTTAAATACTCAGAAA  176

Qy       121 CTGGGATTATAGGCGTGAGCCGCCGCGCCCAGCC----AACATTTTTAAATACTCAGAAA  176
             ||||||| |  |||| |||| ||| ||||| ||||    |||||||| ||||||||||||
Db    132611 CTGGGATTACAGGCGTGAACCACCGCGAGTTAGACCTTTGAGAAATTAAAG 132552

Qy       177 GTAGAGGGAATAGTTATAGTGTACCCCATTTACCCATCACTCAGTTCAACAGCTGGTGA    236
             ||                || || ||  ||||||||||||||||||||||||||||||
Db    132551 ACCTTAAGAGGCTGAGTGCAGTGGCNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 132492

Qy       237 CATATTTATTTCTTCTATACCAGTACCGTACTCTCCCCACTGGGATTATTTTAAGGCAAA    296
             |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    132491 NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 132432

Qy       297 ACCCAGATGACATTTTATCCCTAAATACTTAGATAAAAGGTGTTCTTTGAAAAAAATCAT    356
             ||  |  |||||| |||||| |||||||||||||||||||||||||||||||||||||||
Db    132431 NNNNNNTGGGTAATCCTTCCACTATTAGGGAGGCCAAAGCACGCAGACTACCTTAGTTCAG 132372

Qy       357 AACCTCAGGACCAGCCTGGCCAACATGGTGAAAACCCCGTCGTCTGTTGTACTAAAAATACAAAAAT    416
             |||               |||||||||||||||| ||||||||||||||||||||| |||||||||||
Db    132371 GAGTTCAAGACCAGCCTGGCCAACATGGTGAAAACCCTATCTCTACTAAAAACTACAAAAAT 132312

Qy       417 TAGCTTGGCANGGTCGTGGGCGCACCTGTAATCCCAGCTACTCAGGAGGCTGAGGCAGGAGA    476
             |||| |||||| ||| ||| ||||||||||||||||||||||||||||||||||||||||||
Db    132311 TAGCCAGGCANTGGTGGCGCGTGCCTGTAATCCCAGCTACTCGGGAGGCTGAGGCAGGAGA 132252
```

```
Qy       477 ATCACTTGAACCTCGGGAAGCNGCAGGAGAGGATTGCCAGTGAGCCTGAGATTGCAGTCGAGCCTTGGGCG    536
             ||||||||| ||||||  |||| |||  ||  ||||||||||||||||||||||||||||||||| ||
Db    132251 ATTGCTTGAATCTGGGAGGATGGCAGTGGAGGTTGCCAGTGAGCCGCGAGCTGAGATTGTGCCACTGCACTCCAG 132192

Qy       537 ACAGAGACAGAGAAATGAAACTTCTGTCTCAAAAACAACAACAAAACAAAACAAAACCACTATACATA    596
             ||||| | |||| |||||||  |||||||||||||||||||| |||||||||||||||||||||||||
Db    132191 GCTGAGTGACACAGCGAGACTCATCTCAAAAAGAAAAAGAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 132132

Qy       597 AAAA    600
             ||||
Db    132131 AAAA 132128
```

```
RESULT 4
ACO93822               32289 bp    DNA     linear   PRI 01-MAR-2002
LOCUS       ACO93822        32289 bp    DNA     linear   PRI 01-MAR-2002
DEFINITION  Homo sapiens BAC clone RP11-382L24 from 2, complete sequence.
ACCESSION   ACO93822 AC036160
VERSION     ACO93822.2  GI:18497235
KEYWORDS    HTG.
SOURCE      Homo sapiens.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 32289)
  AUTHORS   Sulston,J.E. and Waterston,R.
  TITLE     Toward a complete human genome sequence
  JOURNAL   Genome Res. 8 (11), 1097-1108 (1998)
  MEDLINE   99063792
   PUBMED   9847074
REFERENCE   2  (bases 1 to 32289)
  AUTHORS   Isak,A., Meyer,R. and Dixon,R.
  TITLE     The sequence of Homo sapiens BAC clone RP11-382L24
  JOURNAL   Unpublished (2001)
REFERENCE   3  (bases 1 to 32289)
  AUTHORS   Waterston,R.H.
  TITLE     Direct Submission
  JOURNAL   Submitted (10-SEP-2001) Genome Sequencing Center, Washington
            University School of Medicine, 4444 Forest Park Parkway, St. Louis,
            MO 63108, USA
REFERENCE   4  (bases 1 to 32289)
  AUTHORS   Waterston,R.H.
  TITLE     Direct Submission
  JOURNAL   Submitted (05-FEB-2002) Genome Sequencing Center, Washington
            University School of Medicine, 4444 Forest Park Parkway, St. Louis,
            MO 63108, USA
REFERENCE   5  (bases 1 to 32289)
  AUTHORS   Waterston,R.
  TITLE     Direct Submission
  JOURNAL   Submitted (08-FEB-2002) Department of Genetics, Washington
            University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
REFERENCE   6  (bases 1 to 32289)
  AUTHORS   Waterston,R.
  TITLE     Direct Submission
  JOURNAL   Submitted (01-MAR-2002) Department of Genetics, Washington
            University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
COMMENT     On Feb 5, 2002 this sequence version replaced gi:15529861.
            -------------- Genome Center
            Center: Washington University Genome Sequencing Center
            Center code: WUGSC
            Web site: http://genome.wustl.edu/gsc
            Contact: sapiens@watson.wustl.edu
            -------------- Summary Statistics
            Center project name: H_NH0382L24
            Drafting Center: WIBR
            --------------
     NOTICE: This sequence may not represent the entire insert of this
            clone.  It may be shorter because we only sequence overlapping
            clone sections once, or longer because we provide a small overlap
            between neighboring data submissions.
```

This sequence was finished as follows unless otherwise noted:
all regions were double stranded, sequenced with an alternate
chemistry, or covered by high quality data (i.e., phred quality >=
30); an attempt was made to resolve all sequencing problems, such
as compressions and repeats; all regions were covered by sequence
from more than one subclone; and the assembly was confirmed by
restriction digest.

MAPPING INFORMATION:
Mapping information for this clone was provided by Dr. John D.
McPherson, Department of Genetics, Washington University, St. Louis
MO. For additional information about the map position of this
sequence, see http://genome.wustl.edu/gsc

SOURCE INFORMATION:
The RPCI-11 human BAC library was made from the blood of one male
donor, as described by Osoegawa,K., Woon,P.Y., Zhao,B., Frengen,E.,
Tateno,M., Catanese,J.J. and de Jong,P.J. (1998) An improved
approach for construction of bacterial artificial chromosome
libraries. Genomics 51:1-8. The clone may be obtained either from
Research Genetics, Inc. (http://www.resgen.com) or Pieter de Jong
and coworkers at http://www.chori.org
VECTOR: pBACe3.6

NEIGHBORING SEQUENCE INFORMATION:
The clone sequenced to the left is RP11-541E12, 2000 bp overlap; the
clone sequenced to the right is RP11-32C8, 2000 bp overlap.
Actual end of this clone is at base position 37821 of RP11-541E12.

Single stranded/single chemistry coverage below phred30, base
position 19583 to 15596.

A trasposon was identified in the vector sequence of the clone
RP11-382L24.

The sequence of AC036160 has been incorporated into AC093822.

FEATURES   Location/Qualifiers
source   1..32289
    /organism="Homo sapiens"
    /db_xref="taxon:9606"
    /chromosome="2"
    /map="2"
    /clone="RP11-382L24"
    /clone_lib="RPCI-11"
repeat_region   3..489
    /rpt_family="ERV1"
repeat_region   607..773
    /rpt_family="MIR"
repeat_region   953..1264
    /rpt_family="Alu"
repeat_region   1236..1264
    /rpt_family="(A)n"
repeat_region   1411..1438
    /rpt_family="(TTTG)n"
repeat_region   1412..1712
    /rpt_family="Alu"
repeat_region   2116..2399
    /rpt_family="Alu"
repeat_region   2376..2399
    /rpt_family="(A)n"
repeat_region   3628..3840
    /rpt_family="MIR"
repeat_region   3913..4006
    /rpt_family="CR1"
repeat_region   4007..4303
    /rpt_family="Alu"
repeat_region   4304..4495
    /rpt_family="Alu"
repeat_region   4500..4595
    /rpt_family="CR1"
repeat_region   4726..4746
    /rpt_family="L2"
repeat_region   4742..4856
    /rpt_family="(T)n"
misc_feature
    /note="similar to EST AA769977 (NID:g2821215) ah7lh12.s1"

repeat_region   5058..5255
    /rpt_family="CR1"
misc_feature   5387..5688
    /note="similar to EST AA769977 (NID:g2821215) ah7lh12.s1"
repeat_region   6124..6229
    /rpt_family="MIR"
repeat_region   6260..6413
    /rpt_family="MIR"
repeat_region   6553..6921
    /rpt_family="ACHobo"
repeat_region   7602..7623
    /rpt_family="ERVL"
repeat_region   7628..7923
    /rpt_family="AT_rich"
repeat_region   7924..8036
    /rpt_family="Alu"
repeat_region   8342..8645
    /rpt_family="Alu"
repeat_region   9085..9443
    /rpt_family="Alu"
misc_feature   9791..10324
    /note="similar to EST AU154949 (NID:g10016470)"
repeat_region   10207..10511
    /rpt_family="Alu"
repeat_region   10585..10885
    /rpt_family="Alu"
repeat_region   10928..11047
    /rpt_family="ERVL"
repeat_region   11048..11432
    /rpt_family="MaLR"
repeat_region   11433..11849
    /rpt_family="MaLR"
repeat_region   11778..11808
    /rpt_family="ERVL"
repeat_region   12065..12446
    /rpt_family="(CA)n"
repeat_region   12472..12537
    /rpt_family="Mariner"
repeat_region   12832..12866
    /rpt_family="AT_rich"
repeat_region   12849..13145
    /rpt_family="Alu"
repeat_region   13431..13509
    /rpt_family="Alu"
repeat_region   13670..13986
    /rpt_family="L2"
repeat_region   13957..13989
    /rpt_family="Alu"
repeat_region   13990..14295
    /rpt_family="(A)n"
repeat_region   14419..14549
    /rpt_family="Alu"
misc_feature   15064..15366
    /rpt_family="MIR"
repeat_region   15441..15573
    /note="similar to EST BF329308 (NID:g11300056)"
repeat_region   15937..16028
    /rpt_family="Alu"
repeat_region   16115..16247
    /rpt_family="MER1_type"
repeat_region   16255..16559
    /rpt_family="Alu"
repeat_region   17089..17130
    /rpt_family="Alu"
repeat_region   17188..17443
    /rpt_family="(CA)n"
repeat_region   17445..17581
    /rpt_family="Alu"
repeat_region   17448..17470
    /rpt_family="L1"

Query Match   23.5%;   Score 234.6;   DB 9;   Length 32289;
Best Local Similarity   65.3%;   Pred. No. 9.4e-55;

RESULT 5

Matches 394; Conservative 0; Mismatches 199; Indels 10; Gaps 3;

```
Qy      3 TGCCACCATGCCTGGCCTAATTTTTCTTTATTTTTAGTTAGAGACCAGGTTTTGCCATGTTGAC   62
          || | || || || |   ||||| |  |||  |||||||| | ||||  ||||||||||||||||
Db  17293 TGTCATCACATGCCCCCACTAATTTTTGTAATTTTAGTAGAGATGAGGTTTTGCCATGTTGCC  17352

Qy     63 CAGGCTGGTCTCGAACTCTTGACCTCAGGTGATCCGCCTGCCTCAGCCTCCCAAAGTGCT    122
          ||||||||||||||||||||||| ||||||||| || ||||||||||||||||||||||||
Db  17353 CAGGCTGGTCTCAAAACTCCTGCACCACGATCTGCACGCCTCGGCCTCCCAAAGTGCT   17412

Qy    123 GGGATTATAGGCGTGAGCCGCGCGCACCCAGCC----AACATTTTTTAAATACTGAAAAGT    178
          |||||||||||||||||||||  |||||||||||      ||  ||||   ||||  ||||||||
Db  17413 GGGATTATAGGCGTGAGCCACCACGTGTCGGCCTGGAATATAATTTTAAATTTTATAAAAAGT  17472

Qy    179 AGAGGGAATAGTATAGTGTACCCCATTACCCATCAGTTCAACAGCGTGGTGACA    238
          |  ||| | |  |||   |||||||||||||||||||||||||||| ||| |  ||
Db  17473 TGCAGAGATAGCTCAATGAGCTCCCATATAGTTGCCCCAGCCCTCTCTAATGTTAACA  17532

Qy    239 TATTTATTTCTTCTATACCAGTACCGTACTCTCCCCACTGGGATTATTTTAAGGCAAAAC    298
          | || | ||||||||||||||||||||||| ||||| ||||||||||  ||||  || |
Db  17533 TCTTATTCAACCAGGGTACATTTGTCTAAAACTAAGAAATCAGTGTCCGTTAGCTGGGTACA  17592

Qy    299 CCAGAATGACAATTTATCCCTAAATACTTTAGATAAAGGTGTTCTTTGAAAAAA-----T    353
          |||  | |||| ||  |  |||  |||  ||| | ||||  ||||| |||||||      |
Db  17593 GTGGCTCATGCCTATCATCCCAGCACTTTGGGAGGCCGAGGTGGGTGGATCACCTGAGGT  17652

Qy    354 CATAACCTCAGGACCAGCCTGGCCAACAACATGGTGAAACCCTGTCTGTACTAAAAATACAAA    413
          |||| |  |||  | ||||||||||||||| ||| ||||| |||||||||||||||||||||
Db  17653 CAGGAGTTCAGACCAGCCTGGCCAACATGGTTGAAACCCCGTCTCTACTAAAAATACACA  17712

Qy    414 AATTAGCTTGGCATGGTGGTGTGGGCGCCTGTAATCCCAGCTACTCAGGAGGCTGAGGCAGG    473
          |||||||||||||||||||| |||||||| ||||||||||||||||||| ||||||||||||
Db  17713 AATTAGCTGGATGTGGTGGCGCACGCCTGTAATCCCAGCTACTCAGGAGGCTGAGGCAGG  17772

Qy    474 AGAATCACTTGAATCCGGGAAGCAGAGAGATTGCAGTGAGCTGAGATTGCAGTGAGCCTGG    533
          ||||| |||||||  | ||||||||||| ||||||||| ||| ||||| |||||| |
Db  17773 AGAATCACTAGGACCCGGGAGGCGGAGGTTGCAGTGAGCCAAGATCATACCATTGCACTC  17832

Qy    534 GCGACAGCAGCAGAA-TGAAACTCTGTCTCAAAAAACAAAACAAAACAAAACCACTATA    592
          | ||| | ||| ||  |||  ||  |  ||||||||||||||||||||||||||||||
Db  17833 CAGCCTGGGCGACAAGCGAGACTCCGTCTCAAAAAAGAAAAGAAAAGTAATCAA  17892

Qy    593 CAT CAT    595
          ||| |||
Db  17893 TAT  17895
```

REFERENCE   1 (bases 1 to 186625)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C. and Lander,E.
  TITLE   Homo sapiens chromosome 17, clone hRPC.1029_k.10
  JOURNAL   Unpublished
REFERENCE   2 (bases 1 to 186625)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C., Lander,E., Allen,N., Anderson,M.,
            Baker,J., Baldwin,J., Barna,N., Beckerly,R., Benn,J., Brown,A.,
            Castle,A., Cerny,J., Colangelo,M., Collins,S., Collymore,A.,
            Cooke,P., DeArellano,K., Depayre,E., Devon,K., Dewar,K.,
            Donelan,L., Doyle,M., Ferreira,P., FitzHugh,W., Forrest,C.,
            Funke,R., Gage,D., Galagan,J., Gardyna,S., Gilbert,D., Grant,G.,
            Hagos,B., Heaford,A., Horton,L., Howland,J.C., Jones,C., Kann,L.,
            Karatas,A., Lehoczky,J., Lieu,C., Locke,K., Macdonald,P.,
            Marquis,N., McEwan,P., McGurk,A., McKernan,K., McLaughlin,J.,
            Meldrim,J., Molla,M., Morris,W., Morrow,J., Mychaleckyj,J.,
            Naylor,J., Niloff,M., O'Connor,T., O'Donnell,P., Pavlin,B.,
            Peterson,K., Pollara,V., Riley,R., Roberts,D., Roy,A., Severy,P.,
            Stange-Thomann,N., Stojanovic,N., Stone,C., Subramanian,A.,
            Tesfaye,S., Torruella-Miller,I., Vassilev,H., Vo,A., Wagner,A.,
            Wheeler,J., Wu,X., Wyman,D., Ye,W.J. and Zody,M.
  TITLE   Direct Submission
  JOURNAL   Submitted (02-MAR-1999) Whitehead Institute/MIT Center for Genome
            Research, 320 Charles Street, Cambridge, MA 02141, USA
REFERENCE   3 (bases 1 to 186625)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C., Lander,E., Allen,N., Anderson,M.,
            Baker,J., Baldwin,J., Barna,N., Beckerly,R., Benn,J., Brown,A.,
            Castle,A., Cerny,J., Colangelo,M., Collins,S., Collymore,A.,
            Cooke,P., DeArellano,K., Depayre,E., Devon,K., Dewar,K.,
            Donelan,L., Doyle,M., Ferreira,P., FitzHugh,W., Forrest,C.,
            Funke,R., Gage,D., Galagan,J., Gardyna,S., Gilbert,D., Grant,G.,
            Hagos,B., Heaford,A., Horton,L., Howland,J.C., Jones,C., Kann,L.,
            Karatas,A., Lehoczky,J., Lieu,C., Locke,K., Macdonald,P.,
            Marquis,N., McEwan,P., McGurk,A., McKernan,K., McLaughlin,J.,
            Meldrim,J., Molla,M., Morris,W., Morrow,J., Mychaleckyj,J.,
            Naylor,J., Niloff,M., O'Connor,T., O'Donnell,P., Pavlin,B.,
            Peterson,K., Pollara,V., Riley,R., Roberts,D., Roy,A., Severy,P.,
            Stange-Thomann,N., Stojanovic,N., Stone,C., Subramanian,A.,
            Tesfaye,S., Torruella-Miller,I., Vassilev,H. Vo.A., Wagner,A.,
            Wheeler,J., Wu.X., Wyman,D., Ye,W.J. and Zody,M.
  TITLE   Direct Submission
  JOURNAL   Submitted (02-FEB-1999) Whitehead Institute/MIT Center for Genome
            Research, 320 Charles Street, Cambridge, MA 02141, USA
REFERENCE   4 (bases 1 to 186625)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C., Lander,E., Allen,N., Anderson,M.,
            Baker,J., Baldwin,J., Barna,N., Beckerly,R., Benn,J., Brown,A.,
            Castle,A., Cerny,J., Colangelo,M., Collins,S., Collymore,A.,
            Cooke,P., DeArellano,K., Depayre,E., Devon,K., Dewar,K.,
            Donelan,L., Doyle,M., Ferreira,P., FitzHugh,W., Forrest,C.,
            Funke,R., Gage,D., Galagan,J., Gardyna,S., Gilbert,D., Grant,G.,
            Hagos,B., Heaford,A., Horton,L., Howland,J.C., Jones,C., Kann,L.,
            Karatas,A., Lehoczky,J., Lieu,C., Locke,K., Macdonald,P.,
            Marquis,N., McEwan,P., McGurk,A., McKernan,K., McLaughlin,J.,
            Meldrim,J., Molla,M., Morris,W., Morrow,J., Mychaleckyj,J.,
            Naylor,J., Niloff,M., O'Connor,T., O'Donnell,P., Pavlin,B.,
            Peterson,K., Pollara,V., Riley,R., Roberts,D., Roy,A., Severy,P.,
            Stange-Thomann,N., Stojanovic,N., Stone,C., Subramanian,A.,
            Tesfaye,S., Torruella-Miller,I., Vassilev,H., Vo.A., Wagner,A.,
            Wheeler,J., Wu.X., Wyman,D., Ye,W.J. and Zody,M.
  TITLE   Direct Submission
  JOURNAL   Submitted (09-NOV-2000) Whitehead Institute/MIT Center for Genome
            Research, 320 Charles Street, Cambridge, MA 02141, USA

COMMENT   On Nov 9, 2000 this sequence version replaced gi:4314418.
            All repeats were identified using RepeatMasker: Smit, A.F.A. &
            Green, P. (1996-1997)
            http://ftp.genome.washington.edu/RM/RepeatMasker.html.

FEATURES   Location/Qualifiers
     source   1..186625
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="17"
                     /map="17"
                     /clone="hRPC.1029_k_10"
                     /clone_lib="Peter de Jong/ human PAC library"
     repeat_region   complement(295..406)
                     /rpt_family="L2"
     repeat_region   2012..2103
                     /rpt_family="MIR"
     repeat_region   2207..2247
                     /rpt_family="(TTG)n"
     repeat_region   complement(2249..2529)
                     /rpt_family="Alu"
     repeat_region   complement(3306..3492)
                     /rpt_family="MIR"
     repeat_region   3609..3886
                     /rpt_family="AluJo"
     repeat_region   3895..3947
                     /rpt_family="(CAAA)n"

```
repeat_region    5272..5407
                 /rpt_family="MIR"
repeat_region    6224..6339
                 /rpt_family="MIR"
repeat_region    6342..6438
                 /rpt_family="(ATGGTG)n"
repeat_region    6570..6852
                 /rpt_family="(TGG)n"
repeat_region    6853..6882
                 /rpt_family="AluSx"
repeat_region    6906..7023
                 /rpt_family="(TAAAA)n"
repeat_region    complement(7061..7222)
                 /rpt_family="L2"
repeat_region    7924..7951
                 /rpt_family="MIR"
repeat_region    complement(7952..8224)
                 /rpt_family="(TTTG)n"
repeat_region    8706..8844
                 /rpt_family="AluSx"
repeat_region    9522..9579
                 /rpt_family="(TG)n"
repeat_region    complement(10383..10497)
                 /rpt_family="GC-rich"
repeat_region    complement(11304..11426)
                 /rpt_family="GA-rich"
repeat_region    complement(12270..12515)
                 /rpt_family="MIR"
repeat_region    complement(13244..13433)
                 /rpt_family="MIR"
repeat_region    complement(13717..13837)
                 /rpt_family="MIR"
repeat_region    complement(13882..13904)
                 /rpt_family="AluY"
repeat_region    13905..13927
                 /rpt_family="(TTTA)n"
repeat_region    complement(13928..14209)
                 /rpt_family="AluY"
repeat_region    complement(14690..14937)
                 /rpt_family="L2"
repeat_region    complement(15871..16049)
                 /rpt_family="L2"
repeat_region    16541..16596
                 /rpt_family="MER91A"
repeat_region    17136..17229
                 /rpt_family="G-rich"
repeat_region    complement(17230..17546)
                 /rpt_family="(TTCC)n"
repeat_region    17547..17552
                 /rpt_family="AluSx"
repeat_region    18320..18387
                 /rpt_family="(TTCC)n"
repeat_region    complement(18781..18983)
                 /rpt_family="CT-rich"
repeat_region    19057..19095
                 /rpt_family="MIR"
repeat_region    complement(19096..19407)
                 /rpt_family="CT-rich"
repeat_region    19408..19469
                 /rpt_family="AluSx"
repeat_region    19916..1997
                 /rpt_family="CT-rich"
repeat_region    complement(24453..24599)
                 /rpt_family="MER91A"
repeat_region    25188..25221
                 /rpt_family="MIR"
repeat_region    complement(25798..25885)
                 /rpt_family="(TCCTG)n"
repeat_region    28378..28406
                 /rpt_family="L1MC/D"
repeat_region    complement(30476..30926)
                 /rpt_family="(CA)n"
repeat_region    complement(30927..31244)
                 /rpt_family="L2"
```

```
repeat_region    complement(31245..31424)
                 /rpt_family="AluJb"
repeat_region    31487..31656
                 /rpt_family="L2"
repeat_region    31657..31967
                 /rpt_family="THE1C"
repeat_region    31968..32115
                 /rpt_family="AluSx"
repeat_region    32116..32417
                 /rpt_family="THE1C"
repeat_region    32418..32484
                 /rpt_family="AluSg"
repeat_region    32621..32644
                 /rpt_family="THE1C"
repeat_region    complement(32645..32766)
                 /rpt_family="(TTTTG)n"
repeat_region    complement(32789..32900)
                 /rpt_family="FLAM_C"
repeat_region    32945..33341
                 /rpt_family="L2"
repeat_region    complement(33357..33560)
                 /rpt_family="L2"
repeat_region    complement(33561..33690)
                 /rpt_family="MER33"
repeat_region    complement(33691..33808)
                 /rpt_family="FLAM_C"
repeat_region    complement(35999..36450)
                 /rpt_family="MER33"
repeat_region    complement(36458..36714)
                 /rpt_family="L2"
```

```
Query Match   Best Local Similarity   23.0%;   Score 230;   DB 9;   Length 186625;
                                      63.0%;   Pred. No. 2.2e-53;
              Conservative   0;   Mismatches 210;   Indels 12;   Gaps 1;
Matches  378;

Qy        1  CCTGCCACCATGCCTGGCTAATTTTCTTATTTTTAGTAGAGACGAGGTTTTGCCATGTTG    60
             || |||| | ||||| || | ||||||| || |||||||||||||||| |||| || |||||
Db   183557  CCCGCCCATCATGCCCCCGGATAATAATTTTTGCATTTTTAGTAGAGACAGGGTTTCACTATGTTG  183616

Qy       61  ACCAGGCTGGTCTCGAAACTCTTGACCTCAGGTGATCCGCGCCTGCCTCAGCCTCCCAAAGTG   120
             ||||||||||| | | |||||||  ||||||||||||| |  ||| ||||  |||||||||||||
Db   183617  GCCAGGCTGGTCGTTGAAACTCCCGACCTCAGGTGATCCACCCGTCTCAGCCTCCCAAAGTC  183676

Qy      121  CTGGGGATTATAGGCGTGAGCCGCCGCCACCCAGCCAACATTTTTTAAATACTGAAAAGTAG   180
             |||||| ||||| ||| |||| | ||  |  ||||||   |||| ||  ||| | ||||| ||
Db   183677  CTGGGATTACAGGCGGTGAGCCACCACGCCCAGCCTAAGATATTTTTATATTAAAAAATTT  183736

Qy      181  AGGGAATAGTTATATAGTGTACCCCATTTACCCCATCACTCAGTTTCAACAGCTGGTGACATA   240
             |  |                       |||  | | |                    |
Db   183737  AAAGAA------------ACACCCAAGAATGATCAATAAAAAATAAATTAAAAAAAAAAA  183784

Qy      241  TTTATTTCTTCTATACCAGTACCGTACTCTCCCACTGGGGATTATTTTAAGGCAAAACCC   300
             ||| |  |   |||      | |  |  ||||||||||  ||||||||| ||||||||||
Db   183785  AAAGCAAGTTGTTAGTTACGGAATATTAAAAAAAAATTAAGACCCAGGCATGGTGGC  183844

Qy      301  AGATGACATTTTATCCCTAAATACTTTAGATAAAGGTGTTCTTTGAAAAAAACATAACC   360
             |   |||| || ||||     |  |||||  |  ||| || ||| |||||| || ||
Db   183845  TCATGCCTGTAAATCCCAGGACTTTGGGGAGGCTGAGGTGGGCGGATCACAAGGTCAGGAGA  183904

Qy      361  TCAGGACCAGCCTGGCCAACATGGTGAAACCCTGTCTGTACTAAAAATACAAAAAATTAGC   420
             ||  ||||| ||||||||||||||||||||||  ||| |||||||||||||||||||||||
Db   183905  TCGAGACCATCCTGGCCAACATGGTGAAACCCGTCTCTACTAAAAATACAAAAAATTAGC  183964

Qy      421  TTTGGCATGGTGCGTGGGCACCTGTAATCCCAGCTACTCAGGAGGCTGAGGCAGGAGAATCA   480
             |  | | ||| ||||||| |||||| ||||||||||||| ||||||||||||||||||||
Db   183965  TGGGCGTGGTGGTGCGCCTGTAGTCCCAGCTACTCGGGAGGCTGAGGCAGGAGAATGG  184024

Qy      481  CTTGAATCCGGGAAGCAGAGAATTGCAGTGAGCTGAGATTGCAGTGCAGCTGGGCGACAG   540
             ||||   |||||| || ||| |||||||||||| |||||  |  |   ||||| |||||
Db   184025  CTTGGGACCCGGGAGGCGGAGGTTGCAGTGAGCCGAGATCGCACCACTGTACTCCAGCCTG  184084

Qy      541  AGACAGAAATGAAAACTCTGTCTCAAAAAACAAACAAAAAAAACCACTATACATAAAAA   600
             |   ||| |||||  ||  |||||||||||||||||||||||| |||||| ||| ||||
Db
```

```
Db 184085 ATGACAGAGCAAGACTCTGCCTCAAAAAAAAAAAAAAAAAAAAAAAAAAATTAAGACAGCATAAAA 184144

RESULT 6
LOCUS       AC015795
DEFINITION  Homo sapiens chromosome 17, clone RP11-613C6, complete sequence.
ACCESSION   AC015795
VERSION     AC015795.20  GI:19698707
KEYWORDS    HTG.
SOURCE      Homo sapiens
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 170862)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C. and Lander,E.
  TITLE     Homo sapiens chromosome 17, clone RP11-613C6
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 170862)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C., Lander,E., Allen,N., Anderson,M.,
            Baldwin,J., Barna,N., Beckerly,R., Boguslavkiy,L., Boukhgalter,B.,
            Brown,A., Castle,A., Colangelo,M., Collins,S., Collymore,A.,
            Cooke,P., DeArellano,K., Dewar,K., Domino,M., Donelan,L., Doyle,M.,
            Ferreira,P., Fitzhugh,W., Forrest,C., Funke,R., Gage,D.,
            Galagan,J., Gardyna,S., Grant,G., Hagos,B., Heaford,A., Horton,L.,
            Howland,J.C., Johnson,R., Jones,C., Kann,L., Karatas,A., Klein,J.,
            Lehoczky,J., Lieu,C., Locke,K., Macdonald,P., Marquis,N.,
            McEwan,P., McKernan,K., McLaughlin,J., Meldrim,J.,
            Morrow,J., Naylor,J., Norman,C.H., O'Connor,T., O'Donnell,P.,
            Peterson,K., Pollara,V., Riley,R., Roy,A., Santos,R., Severy,P.,
            Stange-Thomann,N., Stojanovic,N., Subramanian,A., Talamas,J.,
            Tesfaye,S., Tirrell,A., Vassiliev,H., Vo,A., Wheeler,J., Wu,X.,
            Wyman,D., Ye,W.J., Zimmer,A. and Zody,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (17-NOV-1999) Whitehead Institute/MIT Center for Genome
            Research, 320 Charles Street, Cambridge, MA 02141, USA
REFERENCE   3  (bases 1 to 170862)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C., Lander,E., Ali,A., Allen,N.,
            Anderson,S., Barna,N., Bastien,V., Bloom,T., Boguslavkiy,L.,
            Boukhgalter,B., Brown,A., Camarata,J., Campopiano,A., Chang,J.,
            Chazaro,B., Choepel,Y., Colangelo,M., Collins,S., Collymore,A.,
            Cook,A., Cooke,P., DeArellano,K., Dewar,K., Diaz,J.S., Dodge,S.,
            Faro,S., Ferreira,P., FitzHugh,W., Gage,D., Galagan,J., Gardyna,S.,
            Ginde,S., Gord,S., Goyette,M., Graham,L., Hulme,W., Illey,I.,
            Hagos,B., Horton,L., Hulme,W., Illey,I., Johnson,R., Jones,C.,
            Kamat,A., Karatas,A., Kells,C., LaRocque,R., Lamazares,R.,
            Landers,T., Lehoczky,J., Levine,R., Lindblad-Toh,K., Liu,G.,
            MacLean,C., Macdonald,P., Major,J., Marquis,N., Matthews,C.,
            McCarthy,M., McEwan,P., McKernan,K., Meldrim,J., Meneus,L.,
            Mihova,T., Mlenga,V., Murphy,T., Naylor,J., Nguyen,C., Nicol,R.,
            Norbu,C., Norman,C.H., O'Connor,T., O'Donnell,P., O'Neil,D.,
            Oliver,J., Peterson,K., Phunkhang,P., Pierre,N., Pollara,V.,
            Raymond,C., Retta,R., Rieback,M., Riley,R., Rise,C., Rogov,P.,
            Roman,J., Rosetti,M., Roy,A., Santos,R., Schauer,S., Schupback,R.,
            Seaman,S., Severy,P., Spencer,B., Stange-Thomann,N., Stojanovic,N.,
            Strauss,N., Subramanian,A., Talamas,J., Tesfaye,S., Theodore,J.,
            Topham,K., Travers,M., Travis,N., Trigilio,J., Vassiliev,H.,
            Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J., Young,G.,
            Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (22-MAR-2002) Whitehead Institute/MIT Center for Genome
            Research, 320 Charles Street, Cambridge, MA 02141, USA
REFERENCE   4  (bases 1 to 170862)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C., Lander,E., Ali,A., Allen,N.,
            Anderson,S., Barna,N., Bastien,V., Bloom,T., Boguslavkiy,L.,
            Boukhgalter,B., Brown,A., Camarata,J., Campopiano,A., Chang,J.,
            Chazaro,B., Choepel,Y., Colangelo,M., Collins,S., Collymore,A.,
            Cook,A., Cooke,P., DeArellano,K., Dewar,K., Diaz,J.S., Dodge,S.,
            Faro,S., Ferreira,P., FitzHugh,W., Gage,D., Galagan,J., Gardyna,S.,
            Ginde,S., Gord,S., Goyette,M., Graham,L., Grand-Pierre,N.,
            Hagos,B., Horton,L., Hulme,W., Illey,I., Johnson,R., Jones,C.,
            Kamat,A., Karatas,A., Kells,C., LaRocque,K., Lamazares,R.,
            Landers,T., Lehoczky,J., Levine,R., Lindblad-Toh,K., Liu,G.,
```

```
            MacLean,C., Macdonald,P., Major,J., Marquis,N., Matthews,C.,
            McCarthy,M., McEwan,P., McKernan,K., Meldrim,J., Meneus,L.,
            Mihova,T., Mlenga,V., Murphy,T., Naylor,J., Nguyen,C., Nicol,R.,
            Norbu,C., Norman,C.H., O'Connor,T., O'Donnell,P., O'Neil,D.,
            Oliver,J., Peterson,K., Phunkhang,P., Pierre,N., Pollara,V.,
            Raymond,C., Retta,R., Rieback,M., Riley,R., Rise,C., Rogov,P.,
            Roman,J., Rosetti,M., Roy,A., Santos,R., Schauer,S., Schupback,R.,
            Seaman,S., Severy,P., Spencer,B., Stange-Thomann,N., Stojanovic,N.,
            Strauss,N., Subramanian,A., Talamas,J., Tesfaye,S., Theodore,J.,
            Topham,K., Travers,M., Travis,N., Trigilio,J., Vassiliev,H.,
            Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J., Young,G.,
            Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (24-MAR-2002) Whitehead Institute/MIT Center for Genome
            Research, 320 Charles Street, Cambridge, MA 02141, USA
COMMENT     On Mar 24, 2002 this sequence version replaced gi:19683229.
            All repeats were identified using RepeatMasker:
            Smit, A.F.A. & Green, P. (1996-1997)
            http://ftp.genome.washington.edu/RM/RepeatMasker.html
            ----------------------- Genome Center
            Center: Whitehead Institute/ MIT Center for Genome Research
            Center code: WIBR
            Web site: http://www-seq.wi.mit.edu
            Contact: sequence_submissions@genome.wi.mit.edu
            ----------------------- Project Information
            Center project name: L1619
            Center clone name: 613_C_6
            -----------------------
FEATURES             Location/Qualifiers
     source          1..170862
                     /clone="RP11-613C6"
                     /clone_lib="RPCI-11 Human Male BAC"
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="17"
                     /map="17"
     repeat_region   311..389
                     /rpt_family="(TA)n"
     repeat_region   complement(785..907)
                     /rpt_family="FLAM_A"
     repeat_region   939..987
                     /rpt_family="L2"
     repeat_region   1018..1044
                     /rpt_family="(TTTC)n"
     repeat_region   1057..1336
                     /rpt_family="AluSx"
     repeat_region   1753..2054
                     /rpt_family="AluY"
     repeat_region   complement(2341..2633)
                     /rpt_family="AluSp"
     repeat_region   2381..2385
     unsure          2385
                     /note="<30 qual SNGL region"
     unsure          2448..2452
                     /note="<30 qual SNGL region"
     repeat_region   3422..3736
                     /rpt_family="AluY"
     repeat_region   3839..3925
                     /rpt_family="L1ME4A"
     repeat_region   complement(3926..4218)
                     /rpt_family="AluSg"
     repeat_region   complement(4219..4522)
                     /rpt_family="AluSc"
     repeat_region   4523..4646
                     /rpt_family="L1ME4A"
     repeat_region   4732..4909
                     /rpt_family="HAL1b"
     repeat_region   complement(5489..5662)
                     /rpt_family="L2"
     repeat_region   6174..7128
                     /rpt_family="L2"
     repeat_region   7133..7204
                     /rpt_family="L2"
     repeat_region   7205..7501
                     /rpt_family="L2"
```

```
repeat_region    7502..7686
                 /rpt_family="AluY"
repeat_region    8254..8557
                 /rpt_family="L2"
repeat_region    8921..9221
                 /rpt_family="AluSp"
repeat_region    9426..9595
                 /rpt_family="AluSx"
repeat_region    9596..9930
                 /rpt_family="Charlie1a"
repeat_region    9931..11138
                 /rpt_family="MER1B"
repeat_region    11854..12152
                 /rpt_family="Charlie1a"
repeat_region    12167..12224
                 /rpt_family="AluY"
repeat_region    12231..12524
                 /rpt_family="(CA)n"
repeat_region    12537..12806
                 /rpt_family="AluSx"
repeat_region    12808..12936
                 /rpt_family="L1MB2"
repeat_region    12974..13025
                 /rpt_family="L1MB2"
repeat_region    13077..13375
                 /rpt_family="(CA)n"
repeat_region    13428..13731
                 /rpt_family="AluSx"
repeat_region    13765..13793
                 /rpt_family="AluSq"
repeat_region    13839..14138
                 /rpt_family="AT_rich"
repeat_region    14163..14788
                 /rpt_family="AluSx"
repeat_region    14796..14961
                 /rpt_family="(CA)n"
repeat_region    15238..15305
                 /rpt_family="L1ME1"
complement(15307..15606)
                 /rpt_family="(TA)n"
complement(15607..15896)
                 /rpt_family="AluY"
repeat_region    15901..15925
                 /rpt_family="AluSx"
repeat_region    15945..15972
                 /rpt_family="AT_rich"
repeat_region    15989..16013
                 /rpt_family="AT_rich"
repeat_region    16014..16315
                 /rpt_family="AluSc"
repeat_region    17825..17894
                 /rpt_family="MADE1"
repeat_region    17957..18250
                 /rpt_family="L1MC5"
repeat_region    18256..18390
                 /rpt_family="AluSq/x"
repeat_region    18391..18681
                 /rpt_family="AluY"
```

```
Query Match    22.8%;  Score 228;  DB 9;  Length 170862;
Best Local Similarity  62.8%;
Matches  377;  Conservative  0;  Mismatches 210;  Indels  13;  Gaps  1;
                                    Pred No. 7.9e-53;
```

```
Qy     1 CCTGCCACCATGCCTGGCTAATTTTTCTTTATTTTTAGTGTAGAGACGAGGGTTTGCCATGTTG   60
         || |||| |||  ||||  | ||||||||||||||||||||||||||||||| |||||||||||
Db 15743 CCCGCCATCATCCCGGATAATTTTTCATTTTAGTAGAGACAGGGTTTCACTATGTTG 15802

Qy    61 ACCAGGCTGGTCTCGAACTCTTGACCTCAGGTGATCCGCCTGCCTCAGCCTCCCAAAGTG  120
         ||||||||||| ||||||||||||||||||||| ||| || |||||||||||||||||||
Db 15803 GCCAGGCTGGTCTTGAACTCCCGACCTCAGGTGATCCACCCGCTCACCCTCCCAAAGTC 15862

Qy   121 CTGGGATTATAGGCGTGAGCGCCGCGCACCCACGCCAACACATTTTTTAAATACTGAAAAGTAG  180
         ||||||||||||||||||||||||||||||||||||||||| |||||||||||||||||||
Db 15803 CTGGGATTACAGGCGTGAGCCACCACGCCCAGCCTAAGATAATTTTTATAATTAAAAAATT 15922
```

(Lower section — alignment / database records)

```
Db 15863 CTGGGATTACAGGCGTGAGCGCCACCACGCCCAGCCTAAGATAATTTTTATAATTAAAAAATT 15922
Qy   181 AGGGAATAGTTATAGTGTACCCCATTTACCCATCACACTCAGTTTCAACACGTGGTGACATA  240
Db 15923 AAAGAAAACACCCAAGAATGATCAATAAAATAAAATAATTAAAAAAAGCAAGTTGTA 15982
Qy   241 TTTATTTCTTCTATACCAGTACCGTACTCTCCCCACTGGGATTATTTTAAGGCAAAACCC  300
Db 15983 GTACGGAATATTAAAAAAAAAAT------------TAAGACCAGGCATGGTGGC 16029
Qy   301 AGATGACATTTTATCCCTAAATACTTTAGATAAAGGTGTTCTTTGAAAAAACATAACC  360
Db 16030 TCATGCCTGTAATCCCAGGACTTTGGGAGGCTGAGGTGGGCGGATCACAAGGTCAGGAGA 16089
Qy   361 TCAGGACCAGCGCCTGCCCACACATGGTGAAAAACCCTGTGTCTGTACTAAAAAATACAAAAATTAGC  420
Db 16090 TCGAGACCATCCTGGCCAACATGGTGAAACCCCGTCTCTACTAAAAATACAAAAATTAGC 16149
Qy   421 TTGGCATGGGTGCGTGGGCCACCTGTAATCCCAGCTACTCAGGAGGCTGAGGGAGGAGAATCA  480
Db 16150 TGGGCGTGGTGGTGCGTGCGCCTGTAGTCCCAGCTACTCTCGGGAGGCTGAGGCAGGAGAATGG 16209
Qy   481 CTTGAATCCGGGAAGCAGAGATTGCAGTGAGCTGAGATTGCGAGTGCTGCACGTCGGGGCGACAG  540
Db 16210 CTTGAACCCGGGAGGCGGAGGTTGCAGTGAGCCGAGATCGCACCACTGTACTCCAGCCTG 16269
Qy   541 AGACAGAAATGAAACTCTCTGTCTCAAAAAACAAACAAAACCACTATACATAAAA  600
Db 16270 ATGACAGAGCAAGACTCTGCCTCAAAAAAAAATTAAGACAGCATAA 16329
```

```
RESULT 7
LOCUS       AC025188/c
            AC025188/c
DEFINITION  Homo sapiens chromosome 5 clone CTD-2384B11, complete sequence.
ACCESSION   AC025188
VERSION     AC025188.8  GI:17530707
KEYWORDS    HTG.
SOURCE      Homo sapiens.
  ORGANISM  Homo sapiens.
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 143335)
  AUTHORS   DOE Joint Genome Institute and Stanford Human Genome Center.
  TITLE     Direct Submission
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 143335)
  AUTHORS   DOE Joint Genome Institute.
  TITLE     Direct Submission
  JOURNAL   Submitted (07-MAR-2000) Production Sequencing Facility, DOE Joint
            Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA
REFERENCE   3  (bases 1 to 143335)
  AUTHORS   DOE Joint Genome Institute and Stanford Human Genome Center.
  TITLE     Direct Submission
  JOURNAL   Submitted (23-OCT-2001) DOE Joint Genome Institute and Stanford
            Drive, Walnut Creek, CA 94598, USA
REFERENCE   4  (bases 1 to 143335)
  AUTHORS   DOE Joint Genome Institute and Stanford Human Genome Center.
  TITLE     Direct Submission
  JOURNAL   Submitted (12-DEC-2001) DOE Joint Genome Institute, 2800 Mitchell
            Drive, Walnut Creek, CA 94598, USA
COMMENT     On Dec 12, 2001 this sequence version replaced gi:16328263.
            Draft Sequence Produced by DOE Joint Genome Institute
            www.jgi.doe.gov
            Finishing Completed at Stanford Human Genome Center
            www-shgc.stanford.edu
            Quality: Phrap Quality >=40 99.9% of Sequence;
            Estimated Total Number of Errors is 0.1.
FEATURES             Location/Qualifiers
     source          1..143335
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
```

```
LOCUS       AC025188    143335 bp    DNA    linear    PRI 12-DEC-2001
```

```
RESULT 8
LOCUS       AC020927              110965 bp    DNA     linear   HTG 19-APR-2001
DEFINITION  Homo sapiens chromosome 5 clone CTD-2153E7, WORKING DRAFT SEQUENCE,
            4 ordered pieces.
ACCESSION   AC020927
VERSION     AC020927.4  GI:7711631
KEYWORDS    HTG; HTGS_PHASE2; HTGS_DRAFT; HTGS_ACTIVEFIN.
SOURCE      Homo sapiens
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 110965)
  AUTHORS   DOE Joint Genome Institute.
  TITLE     Sequencing of Human Chromosome 5
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 110965)
  AUTHORS   DOE Joint Genome Institute.
  TITLE     Direct Submission
```

```
  JOURNAL   Submitted (12-JAN-2000) Production Sequencing Facility, DOE Joint
            Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA
            On May 6, 2000 this sequence version replaced gi:7417650.
  COMMENT   ------------Genome Center
            Center: Joint Genome Institute
            Center Code: JGI
            Web site: http://www.jgi.doe.gov
            -----
            Project Information
            Center Project Name: 686404
            Center clone name: CITB-H1_2153E7
            -----
            Summary Statistics
            Consensus quality: 106980 bases at least Q40
            Consensus quality: 109994 bases at least Q30
            Consensus quality: 110455 bases at least Q20
            Estimated insert size: 110000; pulse field gel estimation
            Estimated insert size: 110815; sum-of-contigs estimation
            Quality coverage: 7.03 in Q20 bases; pulse field gel estimation
            Quality coverage: 6.98 in Q20 bases; sum-of-contigs estimation.
            NOTE: This is a 'working draft' sequence. It currently
          * consists of 4 contigs. Gaps between the contigs
          * are represented as runs of N. The order of the pieces
          * is believed to be correct as given, however the sizes
          * of the gaps between them are based on estimates that have
          * provided by the submitter.
          * This sequence will be replaced
          * by the finished sequence as soon as it is available and
          * the accession number will be preserved.
          *       1: contig of 8728 bp in length
          *    8729: gap of unknown length
          *    8828: contig of 2501 bp in length
          *    8829: gap of unknown length
          *   11330: contig of unknown length
          *   11429: gap of unknown length
          *   11430: contig of 29160 bp in length
          *   40590: gap of unknown length
          *   40689: gap of unknown length
          *   40690: contig of 70276 bp in length
          *  110965: contig of 70276 bp in length
FEATURES             Location/Qualifiers
     source          1..110965
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="5"
                     /clone="CTD-2153E7"
                     /clone_lib="CalTech human BAC library D"
```

```
BASE COUNT      32294 a    23208 c    24149 g    31014 t    300 others
ORIGIN

Query Match     22.7%;      Score 226.6;  DB 2;  Length 110965;
Best Local Similarity   64.4%;   Pred. No. 1.9e-52;
Matches   373;  Conservative    0;  Mismatches  199;  Indels   7;  Gaps  2;

Qy        1 CCTGCCACCATGCCTGGCCTAATTTTTTTTCTTAATTTTTAGTAGAGACGAGGTTTTCCATGTTG 120
            || |||||||||||||||||| |||||||||| ||||||||||||||||||||||||||||||||||
Db    99092 GCCAGGCTGGTCTCGAACTCCTCGAACCTCAGGTGATCCGCCTGCCTCCCAAAGTG 99033

Qy      121 CTGGGATTATAGGCGTGAGCCGCCGCACCCAGCCAACATTTTTAAATACTGAAAAGTAG 180
            |||||||||||||||||||||||||||||| |||||| ||| |||||||||||||||
Db    99032 CTGGGATTATAGGCGTGAGCCACCACCACGCCGGCTAAA-----TCCAGACTTTTATGGCA 98978

Qy      181 AGGGAATAGTTATAGTGTACCCCATTTACCCATTCACGACGTTGGTGACATA 240
            |||| ||||||||||| |||||||||||||||||||||||||||||||||
Db    98977 GATGGAAAGGCAAGGGAGCTGGTAAAATGAAGATTCTTCGGATCACCGGTCGTGTCGTGACATA 98918

Qy      241 TTTATTTCTTCTATACCAGTACCGTACGTACTCTCCCCACTGGGATTATATTTTAAGGCAAAACC 300
            ||| || ||| ||||| |||||| ||||||||||||||||||||||| |||| |||||
Db    98917 TGAATCTTTGCCCCAAGAAAATCTGTGTTTTTAACAAGTTTCTCAGGCCAGGTGCCGATGGC 98858

Qy      301 AGATGACAATTTATCCCTAAATACTTTAGATAAAGGT--TGTTCTTTGAAAAAAATCATAA 358
            ||| |||  |||| ||||| |||  ||| |||||  ||||| | ||||||||||||||||
Db    98857 TCATGCCTGTAATCCCAGCACTTTGGGAGGCCGAGGCGGGCGGATCACCTGAGGTCAGGA 98798
```

RESULT 9
AL139396/c

LOCUS   AL139396   178451 bp   DNA   linear   PRI 28-FEB-2002
DEFINITION Human DNA sequence from clone RP11-258C19 on chromosome
            Xp11.21-11.23, complete sequence.
ACCESSION   AL139396
VERSION   AL139396.18 GI:19068211
KEYWORDS   HTG.
SOURCE   human.
  ORGANISM Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1 (bases 1 to 178451)
  AUTHORS   Chapman,J.
  TITLE   Direct Submission
  JOURNAL   Submitted (28-FEB-2002) Wellcome Trust Sanger Institute, Hinxton,
            Cambridgeshire, CB10 1SA, UK. E-mail enquiries: clonerequest@sanger.ac.uk
            humquery@sanger.ac.uk Clone requests: clonerequest@sanger.ac.uk
COMMENT   On Mar 1, 2002 this sequence version replaced gi:10862725.
            During sequence assembly data is compared from overlapping clones.
            Where differences are found these are annotated as variations
            together with a note of the overlapping clone name. Note that the
            variation annotation may not be found in the sequence submission
            corresponding to the overlapping clone, as we submit sequences with
            only a small overlap as described above.
            The following abbreviations are used to associate primary accession
            numbers given in the feature table with their source databases:
            Em:, EMBL; Sw:, SWISSPROT; Tr:, TREMBL; Wp:, WORMPEP; Information
            on the WORMPEP database can be found at
            http://www.sanger.ac.uk/Projects/C_elegans/wormpep This sequence
            was generated from part of bacterial clone contigs of human
            chromosome X, constructed by the Sanger Centre Chromosome X Mapping
            Group. Further information can be found at
            http://www.sanger.ac.uk/HGP/ChrX
            This sequence is the entire insert of clone RP11-258C19 This
            sequence was finished as follows unless otherwise noted: all
            regions were either double-stranded or sequenced with an alternate
            chemistry or covered by high quality data (i.e., phred quality >=
            30); an attempt was made to resolve all sequencing problems, such
            as compressions and repeats; all regions were covered by at least
            one plasmid subclone or more than one M13 subclone; and the
            assembly was confirmed by restriction digest. RP11-258C19 is from
            the library RPCI-11.1 constructed by the group of Pieter de Jong.
            For further details see
            http://www.chori.org/bacpac/home.htm
            VECTOR: pBACe3.6.
FEATURES   Location/Qualifiers
     source   1..178451
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="X"
                     /map="Xp11.21-11.23"
                     /clone_lib="RPCI-11.1"
                     /clone="RP11-258C19"
     misc_feature   178349..178451
                     /note="Single clone region. Assembly confirmed by
                     restriction digest data"

```
Qy    359 CCTCAGGACCAGCCTGGCCCACATAGGGTCGAAACCCTGTCTGTCTGTACTAAAATACAAAAATTA 418
          ||  |||||  ||| ||| ||||||                  |||||||||||||||| ||
Db  98797 GTTCGAGACAACCTACTCAACATGACAAAACCGTGTCTTTACTCAAAATACAAAAACTA 98738

Qy    419 GCTTGGCATGGTCGTGGGCACCTGTAATCCCAGCTACTTCAGGAAGCTGAGGCAGGAGAAT 478
          || ||| ||  ||| ||| ||| |||||||||||||| ||  |||||||||||||||||||
Db  98737 GCCAGGCGTTGGTGGCTCGTAATCCCAGCTACTCGGGAAGGCTGAGGCAGGAGAAT 98678

Qy    479 CACTTGAATCCGGGAGGCAGAGATTGCAGTGAGCTGAGATTGCAGTGAGCTCGAGCCTTGGGGGAC 538
          |||||||| || ||||  ||||| |||||||||||  | |||||||||||  |||||  |||
Db  98677 CGCTTGAACCCAGGAGGCGGAGGTTGCAGTGAGCCGAGATTGCCACTGCACTGCAGCC 98618

Qy    539 AGAGACAGAAATGAAACTCTGTCTCAAAAAACAAACAAAC 577
          | || ||||| |||||||||||||||||||||   ||| |
Db  98617 TGGCGACGAGAGCGACTCTGTCTCAAAAAAAAAGGAGGACC 98579
```

RESULT 10
AC067891

LOCUS   AC067891   167924 bp   DNA   linear   HTG 11-JUL-2001
DEFINITION Homo sapiens chromosome 17 clone RP11-342J16 map 17, WORKING DRAFT
            SEQUENCE, 6 unordered pieces.
ACCESSION   AC067891
VERSION   AC067891.5 GI:14670136
KEYWORDS   HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_FULLTOP.
SOURCE   Homo sapiens
  ORGANISM Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1 (bases 1 to 167924)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C. and Lander,E.
  TITLE   Homo sapiens chromosome 17, clone RP11-342J16
  JOURNAL   Unpublished
REFERENCE   2 (bases 1 to 167924)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C., Lander,E., Abraham,H., Allen,N.,
            Anderson,S., Baldwin,J., Barna,N., Bastien,V., Beda,F.,

BASE COUNT   48101 a 42307 c 40594 g 47449 t
ORIGIN

Query Match   22.7%; Score 226.6; DB 9; Length 178451;
Best Local Similarity 64.9%; Pred. No. 2e-52;
Matches 421; Conservative 0; Mismatches 209; Indels 19; Gaps 5;

```
Qy      1 CCTGCCACCATGCCTGGCCTAATTTCTTTATTTTTAGTAGAGACGAGGTTTGCCATGTTG 60
          ||||||||||  |||||||| |  ||||||| ||||||||||||||||  |||||||||||
Db  99411 CCTGCCACCACCACCTGGCTACTTTTTTGTTAGGTTCTAGTAGAGACAGCAGGTTTCACCATGTTG 99352

Qy     61 ACCAGGCTGGTCTCGAACTCCTGACCTCAGGTGATCC-GCCTGCCTCAGCCTCCCAAAGT 119
          ||||||||||||||||||||||||||||||||||||  ||| |||||||||||||||||||
Db  99351 GCCAGGCTGGTCTCGAACTCCTGACCTCAGGTGATCACCCCACCTCAGCCTCCCAAAGT 99292

Qy    120 GCTGGGATTATAGGCGTGAGCGCGCCCACCCAGCCAACATTTTTTAAATACTGAAAAGTA 179
          |||||| ||| ||||  |||| | |||||||||||  | ||||       |||| ||
Db  99291 GCTGGGTTTACAGGCATGAGCCCACCCAGCCGCCCTGTTT--------TCCATGTTT 99242

Qy    180 CAAGGTACTGAAATATTCAATGCCCATTTTCCCAGCCTTACAGTTCTGTGGGGCATAAA 239
          ||||||||||||||||||||||||||||||||||||||||||||||||||| |||||||
Db  99241 CAAGGTACTGAAATATTCAATGCCCATTTTCCCAGCCTTACAGTTCTGTGGGCATAAA 99182

Qy    240 ATTTATTTCTTCTATACCAGTACCGTACTCTCCCCACTGGGGATTATTGTTAAGGCAAAACC 299
          |  ||| | |   ||| ||  |||||||||||||||||| || |||||  ||||| |||||||
Db  99181 AAAACCT-----TCTCTAACCACCCTCGATAGAACCACACATTTTAGAGGCCAGGCGTGGTG 99128

Qy    300 CAGATCACATTTTATCCCTAAATACATTTAGATAAAACGTGTTGTTCTTTGAAAAAAATCATAAC 359
          |||| |||| |||| |||| | ||||  || |  |||||| | ||||| |||||||||| |||| |
Db  99127 GCTCACGCCTGTAATCCCAGCACTTGGGAGGCCAAGGCGGGCGGATCACACGGGTCAGGAG 99068

Qy    360 CTCAGGACCAGCCCTGGCCAACATGGTGAAAACCCTGTCTGTACTAAAAATACAAAAAATTAG 419
          ||||| ||||||||||||||||||||||||| ||| | |||||||||||||||||||||||||
Db  99067 TTTGAGACCAGCCTGGCCAACATGGTGAAACCCCATCTCTACTAAAAATACAAAAAATTAG 99008

Qy    420 CTTGGCCATGGTGGTGGGCGCCTGTAATCCCAGCTACTCGGGAGGCTGAGGCAGGAGAATC 479
          || || ||||||||  |||||||||||||||||||||||||||||||||||||||||||||
Db  99007 CTGGGCGTGGTGGCGGGCGCCTGTAATCCCAGCTACTCGGGAGGCTGAGGCAGGAGAATC 98948

Qy    480 ACTTGAACCCGGGAGGCAGAGGTTGCAGTGAGCTGAGATCGCGCCACTGCACTCCAGCCTGGGCGACA 539
          ||||||||| ||||||| |||||||||||||||| ||||| | |||||||||||||||||||||||||
Db  98947 GCTTGAACCTGGGAGGCGGAGGTTGCAGTGAGCCGAGATTGTGACCCTGCACTCCAGCCTGGGCGACA 98888

Qy    540 GAGAGAGAAATGAA-ACTCCGTCTCAAAAACAAACAAACAAAAAAAC-CACTATACATAA 597
          | | |||| |||   |||||||||||||||  |||   ||||||||| |||||| |||||
Db  98887 GGGTGACAGAGCAAGACTCCGTCTCAAAAAAAAAAAAAGAATCACATTTTACCTAG 98828

Qy    598 AAATGAACAATGATGGCCACAATAGCCACCAGAGAATTTATAAATACAGA 646
          |||| |||||||  ||||| |||| |||||| ||||||| || ||| |||
Db  98827 AGAGGAAAAATGGTTGGCAGAAGAACACAGGATTATATATAAAGTCAGA 98779
```

```
                                Boguslavkiy,L., Boukhgalter,B., Brown,A., Burkett,G.,
                                Campopiano,A., Castle,A., Choepel,Y., Colangelo,M., Collins,S.,
                                Collymore,A., Cooke,P., DeArellano,K., Dewar,K., Diaz,J.S.,
                                Dodge,S., Domino,M., Doyle,M., Ferreira,P., Fitzhugh,W., Gage,D.,
                                Galagan,J., Gardyna,S., Ginde,S., Goyette,M., Graham,L.,
                                Grand-Pierre,N., Grant,G., Hagos,B., Heaford,A., Horton,L.,
                                Howland,J.C., Illev,I., Jones,C., Jones,C., Kann,L., Karatas,A.,
                                Klein,J., LaRocque,K., Lamazares,R., Landers,T., Lehoczky,J.,
                                Levine,R., Lieu,C., Liu,G., Locke,K., Macdonald,P., Marquis,N.,
                                McCarthy,M., McEwan,P., McKernan,K., McPheeters,R.,
                                Meldrim,J., Meneus,L., Mihova,T., Miranda,C., Mlenga,V., Morrow,J.,
                                Murphy,T., Naylor,J., Norman,C.H., O'Connor,T., O'Donnell,P.,
                                O'Neil,D., Oliver,T.M., Oliver,J., Peterson,K., Pierce,N.,
                                Pisani,C., Pollara,V., Raymond,C., Riley,R., Rogov,P., Rothman,D.,
                                Roy,A., Santos,R., Schauer,S., Severy,P., Spencer,B.,
                                Stange-Thomann,N., Stojanovic,N., Subramanian,A., Talamas,J.,
                                Tesfaye,S., Theodore,J., Tirell,A., Travers,M., Trigllio,J.,
                                Vassiliev,H., Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J.,
                                Young,G., Zainoun,J., Zimmer,A. and Zody,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (27-APR-2000) Whitehead Institute/MIT Center for Genome
            Research, 320 Charles Street, Cambridge, MA 02141, USA
  COMMENT   On Jul 11, 2001 this sequence version replaced gi:11345005.
            All repeats were identified using RepeatMasker:
            http://ftp.genome.washington.edu/RM/RepeatMasker.html
            Smit, A.F.A. & Green, P. (1996-1997)
            --------------------- Genome Center
            Center: Whitehead Institute/ MIT Center for Genome Research
            Center code: WIBR
            Web site: http://www-seq.wi.mit.edu
            Contact: sequence_submissions@genome.wi.mit.edu
            --------------------- Project Information
            Center project name: L7857
            Center clone name: 342_J_16
            --------------------- Summary Statistics
            Sequencing vector: M13; M77815; 28 of reads
            Sequencing vector: Plasmid; n/a; 98% of reads
            Chemistry: Dye-terminator Big Dye; 100% of reads
            Assembly program: Phrap; version 0.960731
            Consensus quality: 166252 bases at least Q40
            Consensus quality: 166888 bases at least Q30
            Consensus quality: 167178 bases at least Q20
            Insert size: 167424; agarose-fp
            Insert size: 170000; agarose-fp
            Quality coverage: 11.6 in Q20 bases; sum-of-contigs
            Quality coverage: 11.8 in Q20.
          * NOTE: This is a 'working draft' sequence. It currently
          * consists of 6 contigs. The true order of the pieces
          * is not known and their order in this sequence record is
          * arbitrary. Gaps between the contigs are represented as
          * runs of N, but the exact sizes of the gaps are unknown.
          * This record will be updated with the finished sequence
          * as soon as it is available and the accession number will
          * be preserved.
          *
          *        1        837: contig of 837 bp in length
          *      838       937: gap of 100 bp
          *      938      7423: contig of 6486 bp in length
          *     7424      7523: gap of 100 bp
          *     7524     15931: contig of 8408 bp in length
          *    15932     16031: gap of 100 bp
          *    16032    102943: contig of 86912 bp in length
          *   102944    103043: gap of 100 bp
          *   103044    151725: contig of 48682 bp in length
          *   151726    151825: gap of 100 bp
          *   151826    167924: contig of 16099 bp in length
FEATURES             Location/Qualifiers
     source          1..167924
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="17"
                     /map="17"
                     /clone="RP11-342J16"
                     /clone_lib="RPCI-11 Human Male BAC"
```

```
     misc_feature    1..837
                     /note="assembly_fragment"
                     /clone_end:SP6
                     vector_side:left"
     misc_feature    938..7423
                     /note="assembly_fragment"
     misc_feature    7524..15931
                     /note="assembly_fragment"
     misc_feature    16032..102943
                     /note="assembly_fragment"
     misc_feature    103044..151725
                     /note="assembly_fragment"
     misc_feature    151826..167924
                     /note="assembly_fragment"
                     /clone_end:T7
                     vector_side:right"

BASE COUNT    47187 a    38503 c    36524 g    45197 t    513 others
ORIGIN

Query Match    22.6%;    Score 226.4;    DB 2;    Length 167924;
Best Local Similarity  62.7%;    Pred. No. 2.2e-52;
Matches  376;    Conservative  0;    Mismatches  211;    Indels  13;    Gaps  1;

Qy          1 CCTGCCACCATGCCTGGCCTAATTTTCTTAATTTCTTTATTTTTAGTAGAGACGGCAGTTTGGCCATGTTG 60
              || | ||||  | ||||| ||  || ||||||    |||||||||||||||||||||||||| ||||||
Db     100844 CCGCCATCATGCCCGGATAATTTTTTGCATTTTTAGTAGTAGAGACAGGGTTTCACTATGTTG 100903

Qy         61 ACCAGGCTGGTCTCGAACTCTTGACCTCAGGTGATCCGCCTGCCTCAGCCTCCCAAGTG 120
              | |||||||||||||| |||| | |||||||| ||| ||| ||||||||||||||||||
Db     100904 GCCAGGCTGGTCTTGAACTCCCGACCTCAGGTGATCACACCGTGCTCAGCCTCCCAAAGTT 100963

Qy        121 CTGGGATTATAGGCGTGAGCCGCCGCGCCCACCGCCAACACATTTTTATTTTATAATTTAG 180
              |||||||||||||||||||||||||||||| |||||| |||| ||||| |||| ||||
Db     100964 TTGGGATTACAGGCGTGAGCCACCACGCCCTAGGATATATTTTTTATTTATAATTTAAAAATT 101023

Qy        181 AGGGAATAGTTATATAGGTGTTACCCCAACTTTACCCATCACTCAGTTTCAACAGCTGGTGTGACATA 240
              |  ||| ||| | ||||||||||| |||||||||||||||||||||| ||||||||||| ||||||
Db     101024 AAAGGAAACACCCAAGAAATGATCAAATAAAAAAAATAAAATTAAAAAAAAAAGCAAGTTGTA 101083

Qy        241 TTTATTTCTTCTATACCAGTACCGTACTCTCCCCACTGGGATTTATTTTTAAGGCAAAAACCC 300
              ||  | |  | || || ||| ||| || | ||||||| ||| |||| | ||||||||||||
Db     101084 GTACGGAATATATTAAAAAAAAAAT--------TAAGACCAGGCATGGTGGC 101130

Qy        301 AGATGACATTTTATCCCTAAAATACTTTAGATAAAGGTGTTCTTTGAAAAAATCATAACC 360
              |  | ||||| || ||||||||| | || |||| |||||  ||||||||||||||||||
Db     101131 TCATGCCTGTAATCCCAGCACTTTGGGAGGCTGAGGTGGGCGGATCACGAAGGTCAGGAGA 101190

Qy        361 TCAGGACCAGCCTGGCCAACATGGTGAAAACCCTGTCTCTGTACTAAAAATACAAAAATTAGC 420
              ||||||||||||||||||||| ||||||||||||| | |||| || ||||| || |||||
Db     101191 TCGAGACCATCCTGGCCAACATGGTGAAAATCCCGTCTCTACTAAAAAATACAAAAATTAGC 101250

Qy        421 TTGGCATGGTGGTCGTGGGCGACCTGTAATCCCAGCTACTCAGGAAGCTGAGGCAGGAGAATCA 480
              |||| | ||| || |||   | |||| ||||||||||||||||||||||| ||||||||| |||
Db     101251 TGGGCGTGGTGGTGGATGTGCGCCTGTAGTCCCAGCTACTCGGGAGGCTGAGGCAGGAGAATGG 101310

Qy        481 CTTGAACTCCGGGAAGCAGAGATTGCAGTGAGCTGAGATTGCAGTCGGACCCTGGGCGACAG 540
              |||| ||||||||||| ||| || |||||||||||||| |||| || || | |||  | ||
Db     101311 CTTGAACCCCGGGAGGCGGAGGTTGCAGTGAGCCGAGATCGCACCACTGTACTCCAGCCTG 101370

Qy        541 AGACAGAAATGAAAACTCTGTCTCAAAAAAACAAAACAAAACAAAACCACTATACATATAAAA 600
              ||||| |||||| |||||||  || |  |||||||||| |||||| ||||| | |||||||
Db     101371 ATGACAGAGCAAGACTCTGCCTCAAAAAAAAAAAATTAAGACAGCATAA 101430
```

ORGANISM   Homo sapiens
           Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
           Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 129043)
AUTHORS   Heilig,R., Petit,J.L., Vico,V., Dasilva,C., Robert,C., Wincker,P.,
           Brottier,P., Cattolico,L., Barbe,V., Pelletier,E., Artiguenave,F.,
           Levy,M., Eckenberg,R., Bruls,T., deBerardinis,V., Cruaud,C.,
           Gyapay,G., Saurin,W. and Weissenbach,J.
TITLE   Sequencing of the human chromosome 14
JOURNAL   Unpublished
REFERENCE 2 (bases 1 to 129043)
AUTHORS   Genoscope.
TITLE   Direct Submission
JOURNAL   Submitted (02-MAY-2001) Genoscope - Centre National de Sequencage :
           BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr :
           Web : www.genoscope.cns.fr)
COMMENT   On May 7, 2001 this sequence version replaced gi:11611133.
           ------------------ Genome Center
           Center: Genoscope / Centre National de Sequencage
           Center code: GS
           Web site: http://www.genoscope.cns.fr/
           Contact: SeqRef@genoscope.cns.fr

           The following BAC sequence is oriented from the T7 to the SP6 end.
           Upstream BAC (overlapping the T7 end) : R-865P7 (AC=AL136001)
           Downstream BAC (overlapping the SP6 end) : R-73M18 ---------------
           Summary Statistics
           Assembly program: Phrap; version 2.0
           Quality coverage: 6.42x in Q20 bases; sum-of-contigs
           ----
           Overall quality chart :
           Range   : bases
           ----
           0   : 2
           1 - 9   : 68
           10 - 19 : 133
           20 - 29 : 531
           30 - 39 : 3561
           40 - 49 : 4838
           50 - 59 : 10020
           60 - 69 : 26649
           70 - 79 : 50082
           80 - 89 : 33159
           90 - 99 :
           Percentage of bases with a quality value >= 40 : 99 %.

FEATURES   Location/Qualifiers
     source   1..129043
                    /organism="Homo sapiens"
                    /db_xref="taxon:9606"
                    /chromosome="14"
                    /clone="C-2134A5"
                    /clone_lib="CalTech-D"
     STS   33953..34155
                    /note="matching EMBL:N94563
                    RHdb:RH75092
                    dbSTS:STS52180
                    Identified using the e-PCR software (G. Schuler)"
     STS   34046..34122
                    /note="matching EMBL:G30197
                    RHdb:RH37966
                    dbSTS:STS17806
                    Identified using the e-PCR software (G. Schuler)"
     STS   102875..102999
                    /note="matching EMBL:R20772
                    RHdb:RH53663
                    RHdb:RH32321
                    dbSTS:STS15940
                    Identified using the e-PCR software (G. Schuler)"

BASE COUNT   33146 a   31488 c   32145 g   32264 t
ORIGIN

Query Match   22.6%; Score 226; DB 9; Length 129043;

Best Local Similarity 64.4%;
Matches 390; Conservative 0; Mismatches 205; Indels 11; Gaps 3;

Qy   4 GCCACCATGCCTGGCTAATTTTCTTTATTTTTAGTAGAGACGGAGGTTTTGCCATGTTGACC   63
           ||||||||| ||| || ||||||| ||||||||||||| |||| |||||||| ||||||||||
Db 24401 GTCACCAAGCCCAGCCTAGTTTTTTGTATTTTTAGCAGGGGTTTCACCATGTTGGCC 24342

Qy   64 AGGGCTGGTCTCGAACTCCTGGGTGATCCGGCCCGCCTCAGCCTCCCAAAGTGCTG 123
           |||||||||||| |||||||| ||||||| ||||| |||||||||||||||||||||
Db 24341 AGGGCTGGTCTTGAACTCCTGACCTCAGGTGATCACCAGCTGCCTCAGCCTCCCAAAGTGCTG 24282

Qy   124 GGATTATAGGCGTGAGCCGCCGCGCCCAGCCCAACATTTTTTAAATACTGAAAAGTAGAGG 183
           ||||| ||||||||||||| ||| ||||||| || ||||||| || ||| | | ||||
Db 24281 GGATTACAGGCGTGAGCCACCACGCCCGGCCTTGCATATTTTTATGTCTGTAPTTCT 24222

Qy   184 GAATAGTTATAGTGTGTACCCCCATTTACCCATCACACTCAGTTTCAACAGCTGTGACATATTTT 243
           | ||| | | | |||| ||||| |||||||||||||||||||||||||||||||||||||
Db 24221 GTGCATATTCTTCTCAGCAGATAATATAAAACTTCCACAGGGGCTCCAGACACTGGGCTT 24162

Qy   244 ATTTCTTCTATACCAGTACCGTACTCTCCCCACTGGGATTATTTTAAGGCAAAAACC--CA 301
           | || || || | |||||||||||||||||||||||||||||||||||||||||| ||
Db 24161 AATAGCTCTTAAAAATATATATATATATATATTTTTTTTTTTCAGAAATATTTTCCAGTA 24102

Qy   302 GATGACATTTTATCCCTAAATACTTTAGATAAAGGTGTTCTTTGAAAAAATCATAACCT 361
           ||| | |||| ||| || || |||| ||| |||||||||||| |||||||||||||| ||
Db 24101 GAGAACTAGAAATCATAAAAAGAAAAAATTAGCCGAGGATCACCTGAGGTCAGGAGTT 24042

Qy   362 CAGGACCACGCCTGGCCAACATGGTGAAACCCCTGTCTCTGTACTAAAAATACAAAAATTAGCT 421
           ||||||||||||||||||||||||||||||||| |||| ||||||||||||||||||||||||
Db 24041 CAAGACCAGCCTGGCCGACATGGCGAAACCCCATCTCTACTAAAAATACAAAAATTAGCC 23982

Qy   422 TGGGCATGGTGGTGGGCGCACCTGTAATCCCAGCTACTCAGGAGGCAGAGAATCAC 481
           |||||| |||||||||||| | ||||||||||||| |||||||||| | |||||||||
Db 23981 GGGCGTTGGTGGTGGGCGCCTCGTAATCCCAGCTGCTCAGGAGGCTGGGGAGGAATCAC 23922

Qy   482 TTGAATCCGGGAGGCAGAGGTTGCAGTGAGCTGAGA---TTGCAGTCGAGCCTGGCGAC 538
           |||||||||||||||||||||||||||||||| |||| ||||| || |||| ||||||
Db 23921 TTGAACCCGGGAGGCAGAGGTTGCAGTGAGCCGAGATCTTTGCACTCACCACGCCGAC 23862

Qy   539 AGAGCAGAAAATGAAACTCTGTCTCAAAAACAAACAAAAAAAACCACTATATCATATAAA 598
           | ||| |||| |||||| | |||||||||| ||||||| | || ||||||
Db 23861 ------AAGAGTGAAACGCCATCTCAAAAAAAAAAAAAAAAAGCCAATAAAA 23808

Qy   599 AATGAA 604
           | |||
Db 23807 ATTGGA 23802

RESULT 12
AC002997/c
LOCUS   AC002997   37784 bp   DNA   linear   PRI 07-OCT-1997
DEFINITION Human DNA from chromosome 19-specific cosmid R26667, genomic
            sequence, complete sequence.
ACCESSION   AC002997
VERSION   AC002997.1 GI:2477519
KEYWORDS   HTG.
SOURCE   Homo sapiens.
ORGANISM   Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1 (bases 1 to 37784)
AUTHORS   Lamerdin,J.E., McCready,P.M., Adamson,A.W., Burkhart-Schultz,K.,
            Garcia,E., Kyle,A., Ramirez,M., Stilwagen,S., Garnes,J.,
            Danganan,L., Bruce,R., Quan,G., Montgomery,M., Ow,D., Kobayashi,A.,
            Olsen,A.O. and Carrano,A.V.
TITLE   Sequence analysis of a 1 Mb region in 19q13.1
JOURNAL   Unpublished
REFERENCE   2 (bases 1 to 37784)
AUTHORS   Lamerdin,J.E.
TITLE   Direct Submission
JOURNAL   Submitted (07-OCT-1997) Human Genome Center, Lawrence Livermore
            National Laboratory, 7000 East Ave., Livermore, CA 94551, USA
COMMENT   Map and sequence oriented from centromere to telomere.

```
FEATURES             Location/Qualifiers
     source          1..37784
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="19"
                     /map="19q13.1 from D19S208 to COX7A1"
                     /clone="R26667"
                     /cell_line="5H12-B"
                     /clone_lib="LL19NCO3 R chromosome 19-specific cosmid
                     library"
                     /note="LL19NCO3 library constructed at LLNL from
                     flow-sorted chromosomes from hybrid 5H12-B, which carries
                     chromosome 19 as its only human chromosome."

R26667 overlaps PAC PC28130 to the left and cosmid F21426 to the
right.

     repeat_region   4..178
                     /rpt_family="Alu"
     repeat_region   complement(788..1056)
                     /rpt_family="Alu"
     repeat_region   2897..3360
                     /rpt_family="Alu"
     repeat_region   complement(3631..3716)
                     /rpt_family="Alu"
     repeat_region   complement(4023..4206)
                     /rpt_family="MLT1"
     misc_feature    complement(4023..4206)
                     /note="predicted exon, program: grail2exons_human_1.3,
                     frame: 1, quality: marginal, score: 49.000"
     repeat_region   complement(4383..4678)
                     /rpt_family="Alu"
     repeat_region   complement(4690..4830)
                     /rpt_family="MLT1"
     repeat_region   5448..5858
                     /rpt_family="Alu"
     repeat_region   6150..6602
                     /rpt_family="Alu"
     repeat_region   7409..7685
                     /rpt_family="Alu"
     repeat_region   8018..8305
                     /rpt_family="Alu"
     repeat_region   complement(8423..8722)
                     /rpt_family="Alu"
     repeat_region   8762..8861
                     /rpt_family="Alu"
     repeat_region   8866..9141
                     /rpt_family="L1"
     repeat_region   9164..9267
                     /rpt_family="L1"
     repeat_region   complement(9590..9868)
                     /rpt_family="Alu"
     repeat_region   9988..10088
                     /rpt_family="Alu"
     misc_feature    10463..10759
                     /note="predicted exon, program: grail2exons_human_1.3,
                     frame: 2, quality: excellent, score: 85.000"
     repeat_region   11150..11683
                     /rpt_family="Alu"
     misc_feature    11808..13086
                     /note="predicted exon, program: grail2exons_human_1.3,
                     frame: 2, quality: good, score: 64.000"
     repeat_region   complement(13957..14248)
                     /rpt_family="Alu"
     misc_feature    14871..15863
                     /note="predicted exon, program: grail2exons_human_1.3,
                     frame: 2, quality: excellent, score: 92.000"
     repeat_region   16106..16395
                     /rpt_family="Alu"
     repeat_region   17263..17693
                     /rpt_family="Alu"
     repeat_region   complement(18118..18401)
                     /rpt_family="Alu"
     repeat_region   18807..19120
                     /rpt_family="Alu"
     repeat_region   19298..19574
                     /rpt_family="Alu"
     repeat_region   complement(20158..20442)
                     /rpt_family="Alu"
     repeat_region   20491..20791
                     /rpt_family="Alu"
     repeat_region   complement(21064..21217)
                     /rpt_family="Alu"
     repeat_region   complement(21492..21693)
                     /rpt_family="Alu"
     repeat_region   21695..22143
                     /rpt_family="Alu"
     repeat_region   22970..23020
                     /rpt_family="Alu"
     repeat_region   complement(24569..24894)
                     /rpt_family="MER5"
     misc_feature    25560..25849
                     /note="DDS similarity to T03155 FB26A8 Fetal brain,
                     Stratagene Homo sapiens cDNA clone FB26A8 3'end (1..327).
                     Score: 581 Identity: 313/327 (95%)"
     repeat_region   complement(26656..26942)
                     /rpt_family="Alu"
     repeat_region   26991..27346
                     /rpt_family="Alu"
     misc_feature    complement(28264..28323)
                     /note="predicted exon, program: grail2exons_human_1.3,
                     frame: 1, quality: good, score: 74.000"
     repeat_region   complement(28345..28635)
                     /rpt_family="Alu"
     misc_feature    complement(28730..28894)
                     /note="predicted exon, program: grail2exons_human_1.3,
                     frame: 0, quality: good, score: 61.000"
     repeat_region   complement(28907..29188)
                     /rpt_family="Alu"
     repeat_region   29333..29428
                     /rpt_family="LTR10"
     repeat_region   complement(29828..30108)
                     /rpt_family="Alu"
     repeat_region   30963..30998
                     /rpt_family="Alu"
     repeat_region   complement(31083..31369)
                     /rpt_family="Alu"
     repeat_region   31373..31539
                     /rpt_family="Alu"
     repeat_region   complement(32052..32709)
                     /rpt_family="Alu"
     repeat_region   33521..33758
                     /rpt_family="Alu"
     repeat_region   complement(34242..34537)
                     /rpt_family="Alu"
     repeat_region   34671..34879
                     /rpt_family="Alu"
     repeat_region   complement(34927..34965)
                     /rpt_family="MER7"
     repeat_region   35188..35456
                     /rpt_family="MER7"
     misc_feature    35545..35669
                     /note="predicted exon, program: grail2exons_human_1.3,
                     frame: 0, quality: excellent, score: 85.000"
     repeat_region   complement(35688..35869)
                     /rpt_family="Alu"
     repeat_region   36471..36532
                     /rpt_family="Alu"
     repeat_region   36565..36855
                     /rpt_family="MER4"
     repeat_region   36861..36994
                     /rpt_family="Alu"
     repeat_region   complement(37144..37245)
                     /rpt_family="MIR"
BASE COUNT    10418 a   8913 c   8981 g   9472 t
ORIGIN

Query Match         22.6%;  Score 225.6;  DB 9;  Length 37784;
Best Local Similarity 66.0%;  Pred No. 3.2e-52;
Matches 391;  Conservative 0;  Mismatches 189;  Indels 12;  Gaps 4;
```

```
Qy     1 CCTGCCACCATGCCTGGCTAATTTTTTTATTTTTTAGTAGAGACGAGGTTTTGCCATGTTG 60
         ||||||||||| |||||||||||||| | ||||||||| || |||| ||||||||||||||
Db  9019 CGTGCCACCACCTGCCTAATTTTTTTGTATTTTTTAATACAGATGGGGTTTTGCCATGTTG 8960

Qy    61 ACCAGGCTGGTCTCGAACTCTTGACCTCAGGTGATCCGCCTGCCTCAGCCTCCCAAAGTG 120
         || ||| ||||||| ||||| |||||||||||||||| |||| ||||||||||||||||||
Db  8959 GCCAGACTGGTCTTGAACTCCTGACCTCAGGTGATCTGCCCGCCTCAGCCTCCCAAAGTG 8900

Qy   121 CTGGGATTATAGGCGTGAGCCACCCGCGCCCAGCCCACCACCTCAGGTGATCCGCCCG 180
         |||||||| ||||||||||||||| | |||| ||
Db  8899 CTGGGATTACAGGCGTGAGCCACCGTGCCCGGCCTTTGTTGTGTGTCTTTAATTTATACTTT 8840

Qy   181 AGGGAATAGTATAGTGTACCCCCATTTACCCATCAGTTCAACAGCTGGTGACATA 240
         |||                        ||||    ||||||||||      ||
Db  8839 CTTGATGATTATTGGGTGTTAAGCATTT---TTTCAAGTACTTATAGAGCCATTTGTATATC 8783

Qy   241 TTTATTTCTTCTATACCAGTACGTACTCTCCCACTGGGATTATTTTAAGGCAAAACC 300
         ||   |  ||          ||  |  ||         ||||    |||
Db  8782 TTCTTTCGTTGGAATGTCTGTTTAAATATTTGCTCATTTAAAAAATTTTAGGCTGGGCAT 8723

Qy   301 AGATGACA---TTTTATCCCTAAATACTTTAGA---TAAAGGTGTTCTTTGAAAAAA 352
         |   |            ||                   |||||  |||| |||||||
Db  8722 GGTGGCTCACACCTGTAATCCTAGCATTTTTGGAGGCTGAAGTGGGTGGATTTGCTTGAG 8663

Qy   353 TCATAACCTCAGGACAGCACGCCTGGCCAACACATGGTGAAAACCCTGTCTGTACTAAAAAATACAA 412
         ||   |  ||||| |||||  ||||||| |||| | |||||| ||||  || | |||||||||||||
Db  8662 TCAGGAGTTTCAGACCAGCCTGGCCACATAGTGAAACCTCTACTAAAAAATACAA 8603

Qy   413 AAATTAGCTTGGCATGGTGGCGTGGGCACCTGTAATCCCAGCTACTCAGGAGGCTGAGGCAG 472
         || |||||| |||| |||||||| ||| |||||||||| |||||||||| ||| |||
Db  8602 AATTTAGCTGGGCGTGGTGGCGGGCGCCTGTAATCCCAGCTACTCAGGAGGCTGAGGCAG 8543

Qy   473 GAGAATCACTTGAATCCGGGAGAGCAGAGATTGCAGTGAGCTGAGATCGCGAGCCTG 532
         ||| ||| ||||||| |  |||| | ||| |||||||||| ||||||
Db  8542 GAGGATCGCTTGAACCCAGGAGGCGGAGGTTGCAGTGAGCCGAGATCATACTAGTGCACT 8483

Qy   533 GGCGACAGAGCGAGA-AATGAAAACTCTGTCTCAAAAAACAAACAAACAAAAA 583
         |||| ||| || |||  |||||| ||||||||||||||||
Db  8482 CCAGCCTGGGCAACAGAGTGAAACTCTGTCTCAAAAAAAAAAAAAAAAAAAAAA 8431
```

```
RESULT 13
AC002511/c
LOCUS       AC002511              98713 bp    DNA     linear   PRI 28-AUG-1997
DEFINITION  Human DNA from chromosome 19-specific PAC PC28130, genomic
            sequence, complete sequence.
ACCESSION   AC002511
VERSION     AC002511.1  GI:2347082
KEYWORDS    HTG.
SOURCE      Homo sapiens.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 98713)
  AUTHORS   Lamerdin,J.E., McCready,P.M., Adamson,A.W., Burkhart-Schultz,K.,
            Garcia,E., Kyle,A., Ramirez,M., Stilwagen,S., Garnes,J.,
            Dananan,L., Bruce,R., Quan,G., Montgomery,M., Ow,D., Kobayashi,A.,
            Olsen,A.O. and Carrano,A.V.
  TITLE     Sequence analysis of a 1Mb region in 19q13.1
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 98713)
  AUTHORS   Lamerdin,J.E.
  TITLE     Direct Submission
  JOURNAL   Submitted (28-AUG-1997) Human Genome Center, Lawrence Livermore
            National Laboratory, 7000 East Ave., Livermore, CA 94551, USA
FEATURES             Location/Qualifiers
     source          1..98713
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="19"
                     /map="overlaps cosmid F16632 to the left and cosmid R26667
                     to the right"
                     /map="19q13.1 from D19S208 to CAPNS"
                     /map="oriented from centromere to telomere"
                     /clone="pc28130"
                     /sex="male"
                     /cell_line="HSF7"
                     /cell_type="fibroblast"
                     /note="see Ioannou et al (1994) Nature Genetics 6: 84-89
                     for more information regarding PAC library construction."
     repeat_region   8..67
                     /rpt_family="Alu"
     repeat_region   complement(69..169)
                     /rpt_family="Tigger2"
     repeat_region   complement(261..730)
                     /rpt_family="Tigger2"
     repeat_region   complement(760..1013)
                     /rpt_family="Alu"
     repeat_region   1057..1274
                     /rpt_family="Alu"
     misc_difference 1257..1258
                     /note="polymorphism in length of Alu tail"
                     /clone="ch19 cosmid F16632"
                     /replace="two fewer A's in cosmid at base 29187-29188"
     repeat_region   complement(1320..1550)
                     /rpt_family="Tigger2"
     repeat_region   1479..1550
                     /rpt_family="MER8"
     misc_difference 2741
                     /standard_name="polymorphism"
                     /clone="ch19 cosmid F16632"
                     /replace="C in cosmid (base 30668) with T in PAC clone"
     misc_difference 2780
                     /standard_name="polymorphism"
                     /clone="ch19 cosmid F16632"
                     /replace="A in cosmid (base 30707) with G in PAC"
     repeat_region   complement(2925..3242)
                     /rpt_family="Alu"
     repeat_region   3453..3740
                     /rpt_family="Alu"
     misc_feature    complement(3990..4054)
                     /note="predicted exon, program: grail2exons_human_1.3,
                     frame: 2, quality: good, score: 70.000"
     misc_feature    complement(4210..4312)
                     /note="predicted exon, program: grail2exons_human_1.3,
                     frame: 2, quality: excellent, score: 92.000"
     misc_feature    4688..7614
                     /standard_name="duplication"
                     /note="duplication flanking
                     10 kb DNA insertion in PAC clone relative to ch19 cosmid
                     F16632
                     locations:
                     4688-7614 on 5'-side of insertion and
                     17160-20068 on 3'-side"
     misc_feature    complement(5499..5778)
                     /note="predicted exon, program: grail2exons_human_1.3,
                     frame: 1, quality: excellent, score: 85.000"
     misc_feature    5566..6063
                     /note="BLASTX similarity to P34976 (290..318); match:
                     0.31, score: 5.8e-13; database searched: nr; TYPE-1
                     ANGIOTENSIN II RECEPTOR (AT1) pir||A48857 AT1 angiotensin
                     II receptor - rabbit >gi|299615 (S59041) AT1"
     misc_feature    5988..6380
                     /note="predicted exon, program: grail2exons_human_1.3,
                     frame: 1, quality: excellent, score: 77.000"
     misc_difference 6225
                     /standard_name="polymorphism"
                     /clone="ch19 cosmid F16632"
                     /replace="G in cosmid (base 34153) with A in PAC"
     misc_feature    6247..6333
                     /note="BLASTX similarity to P34976 (290..215); match: 0.26,
                     score: 5.8e-13; database searched: nr; TYPE-1 ANGIOTENSIN
                     II RECEPTOR (AT1) pir||A48857 AT1 angiotensin
                     II receptor - rabbit >gi|299615 (S59041) AT1"
     misc_difference 6559
```

```
misc_difference   /standard_name="polymorphism"
                  /clone="ch19 cosmid f16632"
                  /replace="G in cosmid f16632 (base 34486) with A in PAC"
       7341
misc_difference   /standard_name="polymorphism"
                  /clone="ch19 cosmid f16632"
                  /replace="T in cosmid f16632 (base 35268) with C in PAC"
       7350
misc_feature      /clone="ch19 cosmid f16632"
                  /replace="C in cosmid f16632 (base 35277) with T in PAC"
       7484..19281
misc_difference   /standard_name="polymorphism?"
                  /clone="ch19 cosmid f16632"
                  /replace="insertion of ~10 kb in PAC clone"
repeat_region     complement(10667..10998)
                  /rpt_family="MER1"
       11157..11490)
repeat_region     complement(11157..11490)
                  /rpt_family="L1"
misc_feature      complement(11607..11717)
                  /note="predicted exon, program: grail2exons_human_1.3,
                  frame: 1, quality: marginal, score: 43.000"
misc_feature      complement(12401..12451)
                  /note="predicted exon, program: grail2exons_human_1.3,
                  frame: 2, quality: excellent, score: 82.000"
repeat_region     12779..12916
                  /rpt_family="MER25"
repeat_region     13127..13375
                  /rpt_family="MER25"
repeat_region     complement(13601..13884)
                  /rpt_family="Alu"
       13900..16647
repeat_region     /rpt_family="L1"
       17927..18967
CDS               /note="hypothetical 38.8 kDa protein similar to G-protein
                  coupled receptors; BLASTx similarity to (U66578) putative
                  G protein-coupled receptor [Homo sapiens]; gi|2231669
                  (U90323) purinergic receptor P2Y9 [Homo sapiens]; Pval=
                  3.7e-20
                  BLASTX similarity to P34976 (33..68); match: 0.33; score:
                  2.4e-13; database searched: nr; TYPE-1 ANGIOTENSIN II
                  RECEPTOR (AT1) pir||A48857 AT1 angiotensin II receptor -
                  rabbit >gi|29615 (S59041) AT1
                  predicted exon, program: grail2exons_human_1.3, frame: 1,
                  quality: excellent, score: 93.000"
                  /codon_start=1
                  /product="PC28130.1"
                  /protein_id="AAB67885.1"
                  /db_xref="GI:2347083"
                  /translation="MDTGPDQOSYFSGNHWFVFSVLLTEFLVGLPLNLLALVFVGKLR
                  CRPAVDVLLNLTRASDLILLFLFPRMVEAANGMHWPLPFILCPLSGFIFTTIYLT
                  ALFLAAVSIERFLSVAHPLWYKTRPRLGQAGLVSVACWLLASAHSVVVIEFSGDIS
                  HSQGTNGTCVLEFWKDQLAILLPVRLEMAVVLFVVPLIITSYCYSRLVWIILGRGGSHR
                  RQRRVAGLVAATLLNFLVCFGPYNVSHVVGYICGESPVWRIYVTLLSTLNSCVDPFVY
                  YFSSSGFQADFHELLRRLCGLWGOWQOESSMELKEOKGGEEQRADRPAERKTSEHSQG
                  CGTGGQVACAEN"
misc_feature      20311..20595
gene              20319..20595
                  /rpt_family="Alu"
misc_difference   21328
                  /standard_name="polymorphism"
                  /clone="ch19 cosmid f16632"
                  /replace="C in cosmid f16632 (base 36784) with T in PAC"
repeat_region     21849..21978
                  /rpt_family="Alu"
repeat_region     21979..22490
                  /rpt_family="Alu"
                  /standard_name="VNTR"
       22166..22270
repeat_region     /rpt_family="ministellite"
       23162..23449
repeat_region     /rpt_family="Alu"
       23544..23665
misc_feature      /note="predicted exon, program: grail2exons_human_1.3,
                  frame: 1, quality: good, score: 61.000"
```

```
repeat_region     24021..24330
                  /rpt_family="Alu"
repeat_region     24520..24800
                  /rpt_family="Alu"
misc_feature      complement(24848..24943)
                  /note="predicted exon, program: grail2exons_human_1.3,
                  frame: 2, quality: excellent, score: 81.000"
repeat_region     25559..25928
                  /rpt_family="Alu"
repeat_region     complement(26232..26521)
                  /rpt_family="Alu"
repeat_region     complement(28145..28453)
                  /rpt_family="Alu"
       28691..30619
gene              /gene="EEF1Ap"
                  /pseudo
misc_feature      28691..28902
                  /gene="EEF1Ap"
                  /note="predicted exon, program: grail2exons_human_1.3,
                  frame: 1, quality: excellent, score: 95.000"
                  /pseudo
                  /gene="EEF1Ap"
                  /note="EF-1-ALPHA-1 pseudogene; BLASTX similarity to HUMAN
                  ELONGATION FACTOR 1-ALPHA 1 (EF-1-ALPHA-1) (EF-TU)
                  pir||EFRB1 translation elongation factor eEF-1 alpha
                  chain - Pval= 1.1e-275"
```

```
Query Match
Best Local Similarity   22.6%;   Score 225.6;   Pred. No. 3.5e-52;   DB 9;   Length 98713;
Matches 391; Conservative  66.0%;   0;  Mismatches 189;  Indels  12;  Gaps  4;

Qy      1 CCTGCCACCATGCCTGGGCTAATTTTTCTTATTTTTTAGTAGAGACGAGGGTTTTGCCATGTTG      60
           | |||||||||                              ||||||||||     ||||||||||
Db  90431 CGTGCCACCACCACCTGGCTAATTTTTGTATTTTTTAATACAGATAGGGGGTTTGCCATGTTG   90372

Qy     61 ACCAGGCTGGTCTCGAACTCTTGACCTCAGGTGATCCGCCGCTTGCCTCAGCCTCCCAAAGTG     120
           ||||   ||| ||||||||||||                 ||||||||||     ||||||||||
Db  90371 GCCAGACTGGTCTTGAACTCCTGACCTCAGGTGATCCGCCCTCAGCGTGCCCAAAGTG       90312

Qy    121 CTGGGATTATAGGCGTGAGCCGCGCCACCGCGCCCAGCCAACATTTTTTAAATACTGAAAAGTAG   180
           |||| |||||||  |||||||||||   |||                ||||||
Db  90311 CTGGGATTACAGGCGTGACCACCGTACCCGGCCGTTTGTTGGTCTTTAATTTATACTTT     90252

Qy    181 AGGGAATAGTTATAGTGTACCCCATTTACCCATCACTCAGTTTCAACAGCTGGTGACATA       240
           |||          ||||          ||||          ||
Db  90251 CTTGATGATTATTGGTGTTAAGCATTT---TTTCAAGTACTTATGAGCCATTGTATATC     90195

Qy    241 TTTATTTTCTTCTATACCAGTACCGTACTCTCCCCACTGGGATTATTTTAAGGCAAAACC      300
           |||||||  | | ||||||||||||           | | ||||||| | ||||
Db  90194 TTCTTTCCGTGGAATGTCTGTTTAAAATATTTTGCTCATTTAAAAAAATTTTAGGCTGGGCAT   90135

Qy    301 AGATGACA----TTTTATCCCTAAAATACTTTAGA----TAAAGGTGTGTCTTTGAAAAAAA    352
           ||||          ||| ||||||||        |||          ||| |||
Db  90134 GGTGGCTCACACCTGTAACCCTAGCATTTTTTGGAGGCTGAGGCGGAGGGTGATTGCTTTGAG   90075

Qy    353 TCATAAACCTTCAGGACCAGCCTGGGCCAACATGGTGAAACCCTGTCTACTAAAAATACA      412
           |||          |||||||||||||||||           | |||||||
Db  90074 TCAGGAGTTTCAGGACCAGCCTGGCCAACAGAGCGAAACCTCATCTACTTACTAAAAATACAA   90015

Qy    413 AAATTAGCTTTGGCATGGTGTCGTGGGCACCTGTAATCCCAGCTACTTCAGGAGGCTGAGGCAG      472
           ||            |||||||      ||| || ||||||||||||||||     ||||
Db  90014 AATTTAGCTGGGCGGGGGTTGGCGGGGCGCCTGTAATCCCAGCTACTTAGGGAGGCTGAGGCAG   89955

Qy    473 GAGAATCACTTGAATCCCGGGAGGCAGAGATTGCAGTGAGCTGAGATTGCGAGTCGAGCCTG      532
           || ||||         ||||||||||||||||||||||||||||     |||||| ||
Db  89954 GAGGATCGCGTTGAACCCAGGAGGCGGAGGTTGCAGTGAGCCGAGATCAGCACATACTACTGCACT   89895

Qy    533 GGGCGACAGAGCAGA-AATGAAAACTCGTCTCAAAAAAACAAAACAAAACAAAAAA      583
           |||          |  |||||||||||||||||||||||     |||||||
Db  89894 CCAGCCTGGGCAACAGAGTGAAAACTCTGTCTCAAAAAAAAAAAAAAAAAAAAA    89843
```

```
LOCUS       AC005863         172203 bp    DNA     linear   PRI 11-NOV-1998
DEFINITION  Homo sapiens chromosome 17, clone hRPK.786_O_4, complete sequence.
ACCESSION   AC005863
VERSION     AC005863.1  GI:3859654
KEYWORDS    HTG.
SOURCE      Homo sapiens
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 172203)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C. and Lander,E.
  TITLE     Homo sapiens chromosome 17, clone hRPK.786_O_4
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 172203)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C., Lander,E., Allen,N., Anderson,M.,
            Baker,J., Baldwin,J., Barna,N., Beckerly,R., Benn,J., Boutwell,C.,
            Brown,A., Castle,A., Cerny,J., Colangelo,M., Collins,S.,
            Collymore,A., Cooke,P., Corliss,D., Depayre,E., Devon,K., Dewar,K.,
            Donelan,L., Ferreira,P., FitzHugh,W., Forrest,C., Funke,R.,
            Gage,D., Gardyna,S., Geralgery,K., Grant,G., Hagos,B., Heaford,A.,
            Herena,L., Horton,L., Howland,J.C., Jacotot,L., Jones,C., Kann,L.,
            Karatas,A., Lehoczky,J., Macdonald,P., Marquis,N., McEwan,P.,
            McGurk,A., McKernan,K., Meldrim,J., Molla,M., Morris,W., Morrow,J.,
            Mychaleckyj,J., Nahf,R., Naylor,J., Niloff,M., O'Connor,T.,
            O'Donnell,P., Pavlin,B., Peterson,K., Riley,R., Roberts,D., Roy,A.,
            Severy,P., Stange-Thomann,N., Stilwell,J., Stojanovic,N., Stone,C.,
            Subramanian,A., Tesfaye,S., Tichovolsky,N., Torruella-Miller,I.,
            Vassiliev,H., Vo,A., Wagner,A., Wheeler,J., Wu,Y., Wyman,D.,
            Ye,W.J., Zhao,J. and Zody,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (24-OCT-1998) Whitehead Institute/MIT Center for Genome
            Research, 320 Charles Street, Cambridge, MA 02141, USA
REFERENCE   3  (bases 1 to 172203)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C., Lander,E., Allen,N., Anderson,M.,
            Baker,J., Baldwin,J., Barna,N., Beckerly,R., Benn,J., Boutwell,C.,
            Brown,A., Castle,A., Cerny,J., Colangelo,M., Collins,S.,
            Collymore,A., Cooke,P., Corliss,D., Depayre,E., Devon,K., Dewar,K.,
            Donelan,L., Ferreira,P., FitzHugh,W., Forrest,C., Funke,R.,
            Gage,D., Gardyna,S., Geralgery,K., Grant,G., Hagos,B., Heaford,A.,
            Herena,L., Horton,L., Howland,J.C., Jacotot,L., Jones,C., Kann,L.,
            Karatas,A., Lehoczky,J., Macdonald,P., Marquis,N., McEwan,P.,
            McGurk,A., McKernan,K., Meldrim,J., Molla,M., Morris,W., Morrow,J.,
            Mychaleckyj,J., Nahf,R., Naylor,J., Niloff,M., O'Connor,T.,
            O'Donnell,P., Pavlin,B., Peterson,K., Riley,R., Roberts,D., Roy,A.,
            Severy,P., Stange-Thomann,N., Stilwell,J., Stojanovic,N., Stone,C.,
            Subramanian,A., Tesfaye,S., Tichovolsky,N., Torruella-Miller,I.,
            Vassiliev,H., Vo,A., Wagner,A., Wheeler,J., Wu,Y., Wyman,D.,
            Ye,W.J., Zhao,J. and Zody,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (11-NOV-1998) Whitehead Institute/MIT Center for Genome
            Research, 320 Charles Street, Cambridge, MA 02141, USA
COMMENT     On Nov 11, 1998 this sequence version replaced gi:3858905.
            All repeats were identified using RepeatMasker: Smit, A.F.A. &
            Green, P. (1996-1997)
            http://ftp.genome.washington.edu/RM/RepeatMasker.html.
FEATURES             Location/Qualifiers
     source          1..172203
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="17"
                     /map="17"
                     /clone="hRPK.786_O_4"
                     /clone_lib="RPCI-11 human BAC library"
     repeat_region   complement(167..360)
                     /rpt_family="MIR"
     repeat_region   1333..1519
                     /rpt_family="MIR"
     repeat_region   1694..1793
                     /rpt_family="HAL1"
     repeat_region   complement(3010..3053)
                     /rpt_family="MIR"
     repeat_region   complement(4785..4883)
                     /rpt_family="MIR"
     repeat_region   5008..5053
                     /rpt_family="AluJo/FLAM"
     repeat_region   complement(6394..6906)
                     /rpt_family="purine-rich"
     repeat_region   complement(6907..7213)
                     /rpt_family="L1MB5"
     repeat_region   complement(7214..8089)
                     /rpt_family="AluSg"
     repeat_region   complement(8090..8602)
                     /rpt_family="L1MB5"
     repeat_region   complement(8603..8786)
                     /rpt_family="L1MA7"
     repeat_region   8787..8915
                     /rpt_family="L1MB5"
     repeat_region   8916..9213
                     /rpt_family="AluSx"
     repeat_region   complement(9214..10041)
                     /rpt_family="L1MB5"
     repeat_region   complement(10029..11723)
                     /rpt_family="L1M4c"
     repeat_region   complement(11907..12179)
                     /rpt_family="L1M4"
     repeat_region   12578..12805
                     /rpt_family="L1ME3A"
     repeat_region   13494..13705
                     /rpt_family="L1MB6"
     repeat_region   complement(13792..13824)
                     /rpt_family="(TAAA)n"
     repeat_region   complement(14278..14355)
                     /rpt_family="L2"
     repeat_region   14505..14828
                     /rpt_family="MER93"
     repeat_region   16439..16466
                     /rpt_family="AT_rich"
     repeat_region   16761..16800
                     /rpt_family="AT_rich"
     repeat_region   16802..16963
                     /rpt_family="MSTA"
     repeat_region   complement(16996..17072)
                     /rpt_family="L1PA9"
     repeat_region   17072..17261
                     /rpt_family="L1P"
     repeat_region   17262..17485
                     /rpt_family="MSTA"
     complement(17877..18180)
                     /rpt_family="AluSg"
     repeat_region   21880..21900
                     /rpt_family="AT_rich"
     repeat_region   22911..22938
                     /rpt_family="AluYb8"
     repeat_region   25215..25241
                     /rpt_family="AT_rich"
     repeat_region   25393..25481
                     /rpt_family="AT_rich"
     repeat_region   25722..25916
                     /rpt_family="L1M4"
     complement(26329..26391)
                     /rpt_family="purine-rich"
     repeat_region   27363..28936
                     /rpt_family="L1PB3"
     complement(29996..30194)
                     /rpt_family="L1MC5"
     complement(30210..30333)
                     /rpt_family="L1M"
     repeat_region   30668..30747
                     /rpt_family="AT_rich"
     repeat_region   30867..31162
                     /rpt_family="AT_rich"
```

```
repeat_region   /rpt_family="AluSx"
                complement(31436..31490)
repeat_region   /rpt_family="L1M4"
                complement(31547..31839)
repeat_region   /rpt_family="L1PA13"
                complement(32023..33179)
repeat_region   /rpt_family="L1MEc"
                33180..33206
repeat_region   /rpt_family="(TA)n"
                complement(33207..33483)
repeat_region   /rpt_family="L1MEc"
                33893..34207
repeat_region   /rpt_family="L1MEc"
                34545..34620
repeat_region   /rpt_family="L1MEc"
                34655..34680
repeat_region   /rpt_family="AT_rich"
                34728..35524
repeat_region   /rpt_family="L1M4"
                36046..36461
repeat_region   /rpt_family="L1M4"
                complement(36462..36498)
repeat_region   /rpt_family="(CA)n"
                36499..37042
repeat_region   /rpt_family="L1M4"
                37050..37532
repeat_region   /rpt_family="L1PA15-16"
                39533..39900
repeat_region   /rpt_family="THE1C"
                39901..41110
repeat_region   /rpt_family="THE1C-internal"
                41096..41269
repeat_region   /rpt_family="THE1C-internal"
                41270..41644
repeat_region   /rpt_family="THE1C"
                41720..41800
repeat_region   /rpt_family="L1M2"
                41731..44954
repeat_region   /rpt_family="L1"
                44955..45265
repeat_region   /rpt_family="AluSq"
                complement(45488..45553)
repeat_region   /rpt_family="L1M8"
                complement(45614..46762)
repeat_region   /rpt_family="L1MB5"
                complement(46763..46782)
repeat_region   /rpt_family="(TAAA)n"
                complement(46783..47113)
repeat_region   /rpt_family="L1MB5"
                47114..47397
repeat_region   /rpt_family="AluJo"
                47400..47422
repeat_region   /rpt_family="(TAAA)n"
                complement(47423..47627)
```

```
RESULT 15
AC093229
AC093229/c

LOCUS       AC093229              44183 bp    DNA     linear   HTG 16-AUG-2001
DEFINITION  Homo sapiens chromosome 19 clone CTD-3242E3, WORKING DRAFT
            SEQUENCE, 5 unordered pieces.
ACCESSION   AC093229
VERSION     AC093229.1  GI:15193363
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_ACTIVEFIN.
SOURCE      Homo sapiens
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 44183)
  AUTHORS   DOE Joint Genome Institute.
  TITLE     Sequencing of Human Chromosome 19
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 44183)
  AUTHORS   DOE Joint Genome Institute.
  TITLE     Direct Submission
  JOURNAL   Submitted (16-AUG-2001) Production Sequencing Facility, DOE Joint
            Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA
COMMENT     ------ Genome Center
            Center: Joint Genome Institute
            Center Code: JGI
            Web site: http://www.jgi.doe.gov
            ------
            Project Information
            Center Project Name: 949056, BC014080
            Center clone name: CITB-EL3242E3
            ------
            Summary Statistics
            Consensus quality:    40449 bases at least Q40
            Consensus quality:    41420 bases at least Q30
            Consensus quality:    41761 bases at least Q20
            Estimated insert size: 62550; agarose-fp estimation
            Estimated insert size: 43783; sum-of-contigs estimation
            Quality coverage: 10.3 in Q20 bases; agarose-fp estimation
            Quality coverage: 14.72 in Q20 bases; sum-of-contigs estimation.
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 5 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
```

```
Query Match    22.6%; Score 225.6; DB 9; Length 172203;
Best Local Similarity  63.9%;
Matches   386; Conservative   0; Mismatches 194; Indels  24; Gaps  2;

Qy      1  CCTGCCACCATGCCCGGCTAATTTTCTTATTTTTAGTAGAGACGAGGTTTGCCATGTTG 60
           | |||||||||||||||||||||||  ||||||||||||||||||||||||| |||| |||
Db 125744  CATGCCACCATGCCCGGCTAATTTTTGTATTTTTAGTAGAGACGAGGTTTCGCCATATTG 125803

Qy     61  ACCAGGCTGGTCTCGAACTCCTGACCTCAGGTGATCCGCCTGCCTCCCAAAGTG 120
           ||||||||| |||| ||||||||||||||||||||| ||| | ||| ||||||||
Db 125804  GCCAGGCTGATCTCAAAACTCCTGACCTCAGGTGATCTGCCCGCCTCGGCCTCCCAAAGTG 125863

Qy    121  CTGGGATTATAGGCGTGAGCCGCCGCACCCAGCCACACATTTTTTAAATACTGAAAGTAG 180
           |||||||||||||||||||| ||||  |  ||| |||   |||||||||||| |||| |||
Db 125864  CTGGGATTATAGGCGTGAGCCACCGCGTGCCCGGCCAATTTCTATACCGTTGTCCATACTTT 125923

Qy    181  AGGGAATAGTTATAGTGTACCCCATTACCCATCACTCAGTTTCAACAGCTGGTGACATA 240
           ||          ||        ||       ||         |        |     |
Db 125924  GTTGA----GCCTAAGAATAAAAATGGACAGTTCCATCTGTCACATAAAAACTGTGATCAA 125979

Qy    241  TTTATTTCTTCTATACCAGTACCGTACTCTCCCACTGGGATTATTTTAAGGCAAAACC 300
           | ||||||||| ||||||||||||||||||||||||||||| |||
Db 125980  ATAAATTCGTGTGTCGTGCTAGGCTGGCCACAGTGG------------------C 126019

Qy    301  AGATGACATTTATATCCCTAAATACTTTAGATAAAGGGTGTTCTTTGAAAAAATCATTACC 360
           ||| |  ||   |||| |||| ||  ||||||  |   ||||||||
Db 126020  TCATGCCTGTAATCCCACGCACTTTGGGAGGCCGAGGCGGGCGGATCACGAGGTCATGAGA 126079

Qy    361  TCAGGACCAGCCTGGCCAACATGGTGAAACCCTGTCTCTGTACTAAAAATACAAAATTAGC 420
           || |||||||| |||||||||||||||||||| |||| |||||||||||| | |||||||
Db 126080  TCGAGACCAGCCTGACCAACATGGTGAAACCCCGTGTCTACTAAAAATAGAAAAATTAGC 126139

Qy    421  TTGGCATGGTGGTGCGTGCCTGTAATCCCAGCTACTCGGGAGGCTGAGGCAGGAGAATCA 480
           || ||| ||||| ||| |||||||| |||||||||||||||||||||||||||||||| |
Db 126140  CTGGCGTGGTGGCGGGCGCCTGTAGTCCCAGCTACTCGGGAGGCTGAGGCAGGAGAATTG 126199

Qy    481  CTTGAATCCGGGAAGCAGGAGAATTGCAGTGAGCTGAGATTGCAGTGAGCCTGGGCGACAG 540
           |||||| | ||||||| || |  ||||||||||| ||||||||| | |||| ||||||||
Db 126200  CTTGAACCTGGGAAGCGGAGGTTGCAGTGAGCCGAGATTGCGCCACTGACTCCAGCCTG 126259

Qy    541  AGACAGAGAAATGAAAACTCTGTCTCAAAAACAAAACAAAACAAAACCACTATACATAAAA 600
           || | |||||| ||||  ||||| ||||||||| |||| | |||||     ||  |||||
Db 126260  GTGACGGAGCGAGACTCCGTCTCAAAAAATAAATAAAATAAATAAAATAAAATAAATA 126319

Qy    601  TGAA 604
           | ||
Db 126320  ATAA 126323
```

```
                  *  runs of N, but the exact sizes of the gaps are unknown.
                  *  This record will be updated with the finished sequence
                  *  as soon as it is available and the accession number will
                  *  be preserved.
FEATURES              Location/Qualifiers
    source            1..44183
                      /organism="Homo sapiens"
                      /db_xref="taxon:9606"
                      /chromosome="19"
                      /clone="CTD-3242E3"
                      /clone_lib="Caltech human BAC library D"
                  *           1    1551: contig of 1551 bp in length
                  *        1552    1651: gap of unknown length
                  *        1652    3140: contig of 1489 bp in length
                  *        3141    3240: gap of unknown length
                  *        3241    5838: contig of 2598 bp in length
                  *        5839    5938: gap of unknown length
                  *        5939   13301: contig of 7363 bp in length
                  *       13302   13401: gap of unknown length
                  *       13402   44183: contig of 30782 bp in length.
BASE COUNT      12253 a   9571 c   9152 g  12801 t    406 others
ORIGIN

Query Match             22.5%;   Score 225.2;  DB 2;   Length 44183;
Best Local Similarity   64.9%;   Pred. No. 4.2e-52;
Matches  405;  Conservative   0;  Mismatches 198;  Indels  21;  Gaps  4;

Qy      5 CCACCATGCCCTGGCTAATTTCTTAATTTTTTAGTTAGAGACGGAGGTTTGCCATGTTGACCA 64
          ||||   | || || |||||||| | |||||||||| ||||||||||||||   ||||||| ||
Db  23949 CCACTACACTCAGCTAATTTTGTATTTTTTAGTAGAGACGGAGGTTTCCCCATGTTGGGTCA 23890

Qy     65 GGCTGGGTCTCGAACTCTTGACCTCAGGTGATCCGCCTGCCTCAGCCTCCCAAAGTGCTGG 124
          ||||||||||||||||   | ||||||||||| ||||| | | ||| |  ||||||||||||
Db  23889 GGCTGGGTCTCGAACTCCCGGACCTCAGGTGAATCCGCCCCGTTGCCGCCATCCCAAAGTGCTGG 23830

Qy    125 GATTATAGGCGTGAGCCGCCGCACCCAGCCCAACA----TTTTTTAAATACTGAAAAGTAG 180
          |||  | || ||| | | ||| |||||||| |||  |  |  |  | |||||||||||  ||
Db  23829 GATCACAGTTGTGAAACCACCACGCCCAGCCTGAAGGACCCTTGTATACACTGAAAGCCCGA 23770

Qy    181 AGGGAATAGTTATAGTGTACCCCCATTTACCCATCACTCAGTTCAACAGCTGGGTGACATA 240
           |  || | ||  | ||| | ||| |||  | || ||||||||||  | | | | |  | |
Db  23769 AACAAGAAGGTGGGAGCATCACCCTTGTTCAACCTCAGCCAGGCATGAGGGCATTGAACAAC 23710

Qy    241 TTTATTTCTTCTATACCAGTACCGTACTC----------TCCCCACTGGGATTATTT 287
           | |  ||||| |  |  | ||   | |           ||| ||| | |  |||| |
Db  23709 TCCAGTTCTTCCCTGCTGTTCACGTGTTCATGAAATGGCCAAGAAAACAGTGAGTATATGC 23650

Qy    288 TAAGGCAAAACCCAGATGACATTTTATCCCTAAATACTTTAGATAAAAGG---TGTTCTTT 344
           || |||| | || | || | || | ||| |||| ||||| || | ||||   ||  |
Db  23649 CTGGGCACACAGCAGCTAAATGCCTGTAAATCCCAGCACTTTGAGGAAGCTGAGGCAGGTGGATC 23590

Qy    345 GAAAAAAATCATAAACCTCAGGACCACCAGCCTGGCCAACACATGGTGAAACCCTGTCTGTACTAA 404
           ||   |  |||  | |||||| ||||||||||||||||||| ||| |||||||||||| |||||||
Db  23589 GCCTGAGGTCAGGAGTTCGAGACCAGCCTGGCCAACATGGCGAAACCCTGTCTCTACTAA 23530

Qy    405 AAATACAAAAATTAGCTTGGCATGGTCGTGGGCACCTGTAATCCCAGCTACTCAGGAAGC 464
          ||||||||||||||||  |||| ||| | ||| | |||||||||||||||||||||| ||
Db  23529 AAATACAAAAATTAGCCGGGCGTGGTGGCGGGCGCCTGTAATCCCAGCTACTCAGGAGGC 23470

Qy    465 TGAGGCAGGAGAATCACTTGAATCCGGGAGGCAGAGATTGCAGTGAGCTGAGATTGCAGT 524
          |||||||||||||||||||||| |||||||| | | ||||||||||||||||||||||
Db  23469 TGAGGCAGGAGAATCACTTGAACCCGGGAGGTGGCAGGTTGCAGTGAGCTGAGATTGCACC 23410

Qy    525 CGAGCCTGGGCGACAGAGCGAGA-AATGAAACTCTGTCTCAAAAACAAACAAACAAAAAA 583
          | ||| |||||||||||| |||| |  |||||| ||||||||||| || | | |||
Db  23409 ACTGCACTCCAGCCTGGGCGACAGAGTGAGAACTCCGTCTCAAAAATACATTAATTAATTA 23350

Qy    584 ACCACTATACATAAAAATGAACAA 607
          || | |  |||||||||   |||
Db  23349 ACTAATTTAAAAAAAAAAAAAACCA 23326
```

Search completed: March 29, 2003, 21:04:04

THIS PAGE BLANK (USPTO)

OM protein - protein search, using sw model

GenCore version 5.1.4_p5_4578
Copyright (c) 1993 - 2003 Compugen Ltd.

Run on: March 18, 2003, 15:40:39 ; Search time 17 Seconds
                   (without alignments)
                   643.843 Million cell updates/sec

Title: 1923
Perfect score: 1923
Sequence: 1 MEAKEKQHLLDARPAIRSYT...........QIEDYSEDMKDCQACQGPSD 372

Scoring table: BLOSUM62
                   Gapop 10.0 , Gapext 0.5

Searched: 262574 seqs, 29422922 residues

Total number of hits satisfying chosen parameters: 262574

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
                   Maximum Match 100%
                   Listing first 120 summaries

Database :   Issued_Patents_AA:*
                   1: /cgn2_6/ptodata/2/iaa/5A_COMB.pep:*
                   2: /cgn2_6/ptodata/2/iaa/5B_COMB.pep:*
                   3: /cgn2_6/ptodata/2/iaa/6A_COMB.pep:*
                   4: /cgn2_6/ptodata/2/iaa/6B_COMB.pep:*
                   5: /cgn2_6/ptodata/2/iaa/PCTUS_COMB.pep:*
                   6: /cgn2_6/ptodata/2/iaa/backfiles1.pep:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Match	Query Match Length	DB	ID	Description
1	656.5	34.1	474	4	US-09-461-474-10	Sequence 10, Appli
2	651.5	33.9	448	4	US-09-461-474-8	Sequence 8, Appli
3	649.5	33.8	398	4	US-09-461-474-17	Sequence 17, Appl
4	613.5	31.9	349	4	US-09-461-474-12	Sequence 12, Appl
5	408.5	21.2	342	4	US-09-134-001C-4190	Sequence 4190, Ap
6	385	20.0	359	4	US-09-134-001C-5618	Sequence 5618, Ap
7	342	17.8	157	4	US-09-461-474-14	Sequence 14, Appl
8	210.5	10.9	167	4	US-09-134-001C-2867	Sequence 2867, Ap
9	128	6.7	165	4	US-08-958-207A-520	Sequence 520, App
10	100.5	5.2	154	4	US-09-134-001C-3700	Sequence 3700, App
11	100.5	5.2	370	4	US-08-405-271A-21	Sequence 21, Appl
12	88	4.6	489	4	US-09-134-001C-5241	Sequence 5241, Ap
13	86.5	4.5	766	4	US-09-268-866-2	Sequence 2, Appli
14	85.5	4.4	1212	4	US-09-134-001C-3689	Sequence 3689, Ap
15	85.5	4.4	7257	3	US-09-335-409-5	Sequence 5, Appli
16	85.5	4.4	7257	4	US-09-568-102-5	Sequence 5, Appli
17	85.5	4.4	7257	4	US-09-567-969-5	Sequence 5, Appli
18	85.5	4.4	7257	4	US-09-568-480-5	Sequence 5, Appli
19	85.5	4.4	7257	4	US-09-568-486-5	Sequence 5, Appli
20	85.5	4.4	7257	4	US-09-568-472-5	Sequence 5, Appli
21	85.5	4.4	7257	4	US-09-567-899-5	Sequence 5, Appli
22	85	4.4	387	4	US-08-993-088A-7	Sequence 7, Appli
23	84	4.4	387	4	US-08-993-424B-7	Sequence 7, Appli
24	84	4.4	569	2	US-08-750-723A-2	Sequence 2, Appli
25	84	4.4	569	2	US-09-191-275-2	Sequence 2, Appli
26	83.5	4.3	367	2	US-08-454-549-4	Sequence 4, Appli
27	83.5	4.3	367	3	US-08-454-552-4	Sequence 4, Appli
28	83.5	4.3	367	3	US-08-676-351-3	Sequence 3, Appli
29	83.5	4.3	372	4	US-08-188-275A-4	Sequence 4, Appli
30	83.5	4.3	372	4	US-09-351-198-4	Sequence 4, Appli
31	83.5	4.3	372	4	US-09-113-426-4	Sequence 4, Appli
32	83.5	4.3	518	4	US-09-369-364A-22	Sequence 22, Appli
33	83.5	4.3	930	4	US-09-122-126B-15	Sequence 15, Appl
34	83	4.3	942	3	US-08-560-005-6	Sequence 6, Appli
35	83	4.3	942	3	US-09-418-540-6	Sequence 6, Appli
36	82.5	4.3	485	1	US-08-362-512A-2	Sequence 2, Appli
37	82.5	4.3	485	4	US-09-134-001C-3703	Sequence 3703, Ap
38	82.5	4.3	496	4	US-08-964-939-2	Sequence 2, Appli
39	82.5	4.3	524	4	US-09-134-001C-5457	Sequence 5457, Ap
40	82.5	4.3	591	1	US-08-483-094-2	Sequence 2, Appli
41	82.5	4.3	591	1	US-09-388-774-1	Sequence 1, Appli
42	82	4.3	348	1	US-09-134-001C-4857	Sequence 4857, Ap
43	82	4.3	475	1	US-07-686-591-4	Sequence 4, Appli
44	82	4.3	475	1	US-09-970-715-4	Sequence 4, Appli
45	82	4.3	942	4	US-09-074-579-1	Sequence 1, Appli
46	82	4.3	942	4	US-09-388-774-1	Sequence 1, Appli
47	81.5	4.2	372	4	US-09-149-093A-6	Sequence 6, Appli
48	81.5	4.2	372	2	US-08-911-245-6	Sequence 6, Appli
49	81.5	4.2	372	4	US-09-510-473-6	Sequence 6, Appl
50	80.5	4.2	239	4	US-09-134-001C-4165	Sequence 4165, Ap
51	80.5	4.2	372	6	US-08-430-286A-6	Sequence 6, Appli
52	80.5	4.2	372	4	US-08-120-601B-7	Sequence 7, Appli
53	80.5	4.2	492	4	US-09-413-814-72	Sequence 72, Appl
54	80.5	4.2	504	4	US-08-868-373-6	Sequence 6, Appli
55	80	4.2	345	2	US-08-758-621-14	Sequence 14, Appl
56	80	4.2	345	4	US-09-107-858-14	Sequence 14, Appl
57	80	4.2	808	2	US-09-134-001C-3105	Sequence 3105, Ap
58	79.5	4.1	1279	2	US-08-784-649A-2	Sequence 2, Appli
59	79.5	4.1	1280	3	US-08-583-276-19	Sequence 19, Appl
60	79.5	4.1	1280	6	Patent No. 5206352-4	Sequence 4, Appli
61	79	4.1	339	2	US-08-758-621-2	Sequence 2, Appli
62	79	4.1	339	4	US-08-107-858-2	Sequence 2, Appli
63	79	4.1	1489	6	US-08-669-785-2	Sequence 2, Appli
64	79	4.1	1706	5	Patent No. 5183745-6	Sequence 6, Appl
65	79	4.1	1794	6	US-08-142-648B-7	Sequence 7, Appli
66	78.5	4.1	372	2	US-08-614-377A-7	Sequence 7, Appli
67	78.5	4.1	372	3	US-08-194-290-7	Sequence 7, Appli
68	78.5	4.1	372	4	US-08-292-694A-4	Sequence 4, Appli
69	78.5	4.1	1026	2	US-08-147-592A-4	Sequence 4, Appli
70	78.5	4.1	1026	2	Patent No. 5183745-2	Sequence 2, Appli
71	78.5	4.1	1026	2	US-08-752-447-2	Sequence 2, Appli
72	78.5	4.1	1280	4	US-09-316-167-2	Sequence 2, Appli
73	78.5	4.1	1280	4	US-08-898-976-2	Sequence 2, Appli
74	78.5	4.1	1280	4	US-08-898-976-4	Sequence 4, Appli
75	78	4.1	462	2	US-09-134-001C-2962	Sequence 2962, Ap
76	78	4.1	462	2	US-09-462-136-9	Sequence 9, Appli
77	78	4.1	707	4	US-09-306-998-3	Sequence 3, Appli
78	78	4.1	1296	4	US-09-134-001C-4114	Sequence 4114, Ap
79	77.5	4.1	2037	4	US-08-484-106-10	Sequence 10, Appl
80	77.5	4.0	243	1	US-08-484-105-10	Sequence 10, Appl
81	77.5	4.0	479	1	US-09-335-409-22	Sequence 22, Appli
82	77	4.0	479	1	US-09-568-102-22	Sequence 22, Appli
83	77	4.0	305	4	US-09-567-969-22	Sequence 22, Appli
84	77	4.0	305	4	US-09-568-480-22	Sequence 22, Appli
85	77	4.0	305	4	US-09-568-486-22	Sequence 22, Appli
86	77	4.0	305	4	US-09-568-472-22	Sequence 22, Appli
87	77	4.0	305	4	US-08-805-478-6	Sequence 6, Appli
88	77	4.0	305	4	US-08-802-627A-6	Sequence 6, Appli
89	77	4.0	305	4	US-08-801-238-6	Sequence 6, Appli
90	77	4.0	372	4	US-08-801-228-6	Sequence 6, Appli
91	77	4.0	372	2	US-09-104-296-6	Sequence 6, Appli
92	77	4.0	372	3	US-08-982-493-8	Sequence 8, Appli
93	77	4.0	372	3	US-08-884-681-4	Sequence 4, Appli
94	77	4.0	372	2	US-09-258-643-4	Sequence 4, Appli

101	76.5	4.0	287	3	US-08-549-515-10 Sequence 10, Appl
102	76.5	4.0	372	2	US-08-411-859-2 Sequence 2, Appli
103	76.5	4.0	372	4	US-08-387-707-8 Sequence 8, Appli
104	76.5	4.0	372	4	US-08-405-271A-8 Sequence 8, Appli
105	76	4.0	380	4	US-09-097-889-25 Sequence 25, Appl
106	76	4.0	466	4	US-08-134-001C-3355 Sequence 3355, Ap
107	76	4.0	558	4	US-09-071-035-268 Sequence 268, Appl
108	76	4.0	1638	4	US-09-071-035-258 Sequence 258, App
109	76	4.0	1638	4	US-09-071-035-262 Sequence 262, App
110	76	4.0	1638	4	US-09-071-035-266 Sequence 266, App
111	75.5	3.9	388	2	US-08-759-581B-9 Sequence 9, Appl
112	75.5	3.9	388	4	US-09-304-711-9 Sequence 9, Appl
113	75.5	3.9	388	4	US-09-173-281-9 Sequence 9, Appl
114	75.5	3.9	452	4	US-09-134-001C-3935 Sequence 3935, Ap
115	75.5	3.9	534	2	US-09-031-399-4 Sequence 4, Appl
116	75.5	3.9	534	4	US-09-299-549-4 Sequence 4, Appli
117	75.5	3.9	534	4	US-09-610-417-4 Sequence 4, Appli
118	75.5	3.9	614	4	US-08-291-299-7 Sequence 7, Appli
119	75.5	3.9	614	5	PCT-US95-10579-7 Sequence 7, Appli
120	75	3.9	233	2	US-09-024-848-4 Sequence 4, Appli

ALIGNMENTS

RESULT 1
US-09-461-474-10
Patent No. 6278042
Sequence 10, Application US/09461474

GENERAL INFORMATION:
APPLICANT: Allen, Steve
APPLICANT: Rafalski, Antoni
APPLICANT: Sakai, Hajime
TITLE OF INVENTION: Plant Metal Transporters
FILE REFERENCE: BB1303 US NA
CURRENT APPLICATION NUMBER: US/09/461,474
CURRENT FILING DATE: 1999-12-14
EARLIER APPLICATION NUMBER: 60/112,562
EARLIER FILING DATE: 1998-12-16
NUMBER OF SEQ ID NOS: 17
SOFTWARE: Microsoft Office 97
SEQ ID NO 10
LENGTH: 474
TYPE: PRT
ORGANISM: Oryza sativa
US-09-461-474-10

Query Match   34.18%;   Score 656.5;   DB 4;   Length 474;
Best Local Similarity   31.6%;   Pred. No. 5.2e-65;
Matches 151; Conservative   72; Mismatches 98; Indels 157; Gaps   8;

Qy   17 RSYTGSIWQEGAGWIPLPRPGLDLQAIELAAQSNHHCHAQKGPDSHCD------------ 64
         ||:||| || |:|| || ||
Db   13 RAYTRSL------LPISN-----ARTRHHHLDAGGDDHGDGNGGGGREALLI 54

Qy   65 PK--KGKA------- 74
         || :|
Db   55 PKMDSHNSAPPQIAEVRMDISSSTSVAAGNKVCRGAACDFSDSSNSSKDARERMASMRKL 114

Qy   75 YVASAICLLFMIGEVVGGYLAHSLAVMTDAAHLLTDFASMLISLFSLWMSSRPATKTMNF 134
         | | |||||||||||||||||||||||||||||||||||||||||||||:||||||
Db   115 IIAVILCIIFMAVEVVGGIKANSLAITDAARLLSDVAAFAISLFSLWAAGWEATPQQSY 174

Qy   135 GWQRAEILGALVSVLSIWVVTGVLVLVLAVERLISGDYEIDGGTMLITSGCAVAVNIIMGL 194
         || | |||||| | | |||||||||||||||||||||:|||||||||||||:|||||
Db   175 GFFRIEILGALVSIQLIWLLAGILVYEAIVRLINESGEVQGSLMFAVSAFGLFVNIMAV 234

Qy   195 TL-HQSGHGHSHG--- 209
         | | |||||||| ||
Db   235 LLGHDGHGHGHSHDHDGGSDHDHHHEDQEHGHVHHHHEDGHGNSITVNLHHH 294

Qy   210 ---------------------------------QQEENPSVRAAFIHVIGDFMQSMGVLVA 237
                                          | :| |||:||::|:||||||||::
...

(right column)

Db   295 PGTGHHHHDAEEPLLKSDAGCDSTQSGAKDAKKARRNINVHSAYLHVLGDSIQSIGVMIG 354
Qy   238 AYILLYFKPEYKYVDPICTFEVFSILVLGLTTLTLLRDVLVLVLMEGTPKGVDFTAVRDLLLSV 297
         |:||||||||||| :| | :||:||||||||||||||:||||||||||||:||||||||
Db   355 GAIIWYKPEWKIIDLCICTLIFSVIVLFTTIKMLRNILEVLMESTPREIDATSLENGLRDM 414
Qy   298 EGVEALHSLHIWALTVAQPVLSVHIAIAQNTDAQAVLKTASSRLQGKFHFHTVTIQIE 355
         ||| ||| ||||| ||||| | || | | | |:| :| |||||:|:| |||||
Db   415 DGVVAVHELHIWAITVGKVLLACHVTIFQDADADQOMLDKVIGYIKSEYNISHVTIQIE 472

RESULT 2
US-09-461-474-8
Patent No. 6278042
Sequence 8, Application US/09461474

GENERAL INFORMATION:
APPLICANT: Allen, Steve
APPLICANT: Rafalski, Antoni
APPLICANT: Sakai, Hajime
TITLE OF INVENTION: Plant Metal Transporters
FILE REFERENCE: BB1303 US NA
CURRENT APPLICATION NUMBER: US/09/461,474
CURRENT FILING DATE: 1999-12-14
EARLIER APPLICATION NUMBER: 60/112,562
EARLIER FILING DATE: 1998-12-16
NUMBER OF SEQ ID NOS: 17
SOFTWARE: Microsoft Office 97
SEQ ID NO 8
LENGTH: 448
TYPE: PRT
ORGANISM: zea mays
US-09-461-474-8

Query Match   33.9%;   Score 651.5;   DB 4;   Length 448;
Best Local Similarity   35.3%;   Pred. No. 1.7e-64;
Matches 142; Conservative   71; Mismatches 100; Indels 89; Gaps   6;

Qy   38 LDLQAIELAAQSNHHCHAQKGPDSHCD-------------PKKGKAQRQLYVASAICLLFMI 86
Db   50 MDIAASASGAAGSKFC---KG-AACDFSDASNSSKDARERSASMRKLIVAAVLCVVFMA 104
Qy   87 GEVVGGYLAHSLAVMTDAAHLLTDFASMLISLFSLWMSSRPATKTMNFGWQRAEILGALV 146
Db   105 VEVVGGIKANSLAITDAAHLLSDVAAFAISLFSLWAAGWEATPRQSYGFFRVEILGALV 164
Qy   147 SVLSIWVVTGVLVLVLAVERLISGDYEIDGGTMLITSGCAVAVNIIMGLTL-HQSGHGHSH 205
Db   165 SIQLIWLLAGILVYEAVVRLVGESGDVRGSLMFAVSAFGLAVNVLMAVLLGHDHGHGH 224
Qy   206 G------------------------TTNQQEE-------------------------- 213
Db   225 GHGHSHDHGHGDSDDGHSHHDDEQEOGRVHHHEHGHGGAITVTTHHHHHHHDDVEE 284
Qy   214 ------------NPSVRAAFIHVIGDFMQSMGVLVAAYILLYFKPEYKYVDPI 253
Db   285 ALIKHEGTQSAGRAGKKPRRNINVHSAYLHVLGDSVQSVGVMWYGGAIIWYKPEWKVIDLI 344
Qy   254 CTFVFSILVLGLTTLTLLRDVILVLMEGTPKGVDFTAVRDLLLSVEGVEALHSLHIWALTV 313
Db   345 CTLVFSVVLFTTIRMLRSILEVLMESTPREIDATRLESGLCGMEGVVAVHELHIWAITV 404
Qy   314 AQPVLSVHIAIAQNTDAQAVLKTASSRLQGKHFHFTVTIQIE 355
Db   405 GKVLLACHVTIARDADADEILDKVIGYIKTEYNISHVTIQVE 446

RESULT 3
US-09-461-474-17
Patent No. 6278042
Sequence 17, Application US/09461474

GENERAL INFORMATION:
APPLICANT: Allen, Steve
APPLICANT: Rafalski, Antoni

```
; APPLICANT: Sakai, Hajime
; TITLE OF INVENTION: Plant Metal Transporters
; FILE REFERENCE: BB1303 US NA
; CURRENT APPLICATION NUMBER: US/09/461,474
; CURRENT FILING DATE: 1999-12-14
; EARLIER APPLICATION NUMBER: 60/112,562
; EARLIER FILING DATE: 1998-12-16
; NUMBER OF SEQ ID NOS: 17
; SOFTWARE: Microsoft Office 97
;
; SEQ ID NO 17
; LENGTH: 398
; TYPE: PRT
; ORGANISM: Arabidopsis thaliana

US-09-461-474-17

Query Match         33.8%;
Best Local Similarity   37.9%;   Score 649.5;   DB 4;   Length 398;
Matches 134;   Conservative  71;   Mismatches  88;   Indels  61;   Gaps  4;

Qy    60 DSHCDPKKGKAQRQLYVASAICLLFMIGEVVGGYLAHSLAVMTDAAHLLTDFASMILSLF 119
          |:|                      | ||||| ::|||||||| |||||||||||||||||
Db    47 DAH--ERSASMRKLCIAVVLCLVFMSVEVVGGIKANSLAILTDAAHLLSDVAAFAISLF 103

Qy   120 SLWMSSRPATKTMNFGWQRAEILGALVSVLSLWVVTGVLVYLAVERLISGDYEIDGGTML 179
          |||||||||||||||||||||  | | | ||||  || ||::||||| || ||  |||
Db   104 SLWAAGWEATPROTYGFFRIEILGALVSIQLIWLITGILVEAIIRIVTETSEVNGFLMF 163

Qy   180 ITSGCAVANIIMGLTL------HQSGHGHSHG--------------TIN---------- 209
          ||                    || ||| ||              |||
Db   164 LVAAFGLVVNIIMAVLLGHDHGHSHGHGHGHDHHNHSHGVTVTTHHHHDHEHGHSHG 223

Qy   210 ----------------QQEENPSVRAAFIHVIGDFMQSMGVLVAAYIL----------- 241
                          ||||||||||||||||||||:  ||| |||
Db   224 HGEDKHHAHGDVTEQLLDKSKTQVAAKEKRKRNINLQGAYLHVLGDSIQSVGVMIGGAII 283

Qy   242 YFKPEYKYVDPICTFVFSILVLGTTLTLTILRDVILVLMEGTPKGVDFTAVRDLLLSVEGVE 301
          || :|||  || |||| |||||||||||| |||:::| |||  |||||||||||||| |||
Db   284 WYNPEKKIVDLICTLAFSVIVLGTTINMIRNILEVLMESTPREIDATKIEKGLLEMEEVV 343

Qy   302 ALHSLHIWALTVAQPVLSVHIAIAQNTDAQAVLKTASSRLQGKFHFHTVTIQIE 355
          :||||||||||||:||:|:| |:|||||||||||||||||||||| |||||||
Db   344 AVHELHIWAITVGKVLLACHVNIRPEADADMVLNKVIDYIRREYNISHVTIOIE 397
```

```
US-09-461-474-12

Query Match         31.9%;
Best Local Similarity   37.2%;   Pred. No. 2.2e-60;
Matches 128;   Conservative  62;   Mismatches  81;   Indels  73;   Gaps  3;

Qy    85 MIGEVVGGYLAHSLAVMTDAAHLLTDFASMILSLFSLWMSSRPATKTMNFGWQRAEILGA 144
          |||||| |||:| ||| |||||||||||:|||||||||||||||||||||||  |||||||
Db     5 MTVEVVGGIKANSLAILTDAAHLLSDVASFAISLFSLWAAGWEATPROSYGFFRIEILGA 64
```

```
US-09-134-001C-4190

Query Match         21.2%;
Best Local Similarity   28.4%;   Score 408.5;   DB 4;   Length 342;
Matches 97;   Conservative  80;   Mismatches 144;   Indels  21;   Gaps  7;

Qy   145 LVSVLSIWVVTGVLVYLAVERLISGDYEIDGGTMLIITSGCCAVANIIMGLTL-----HQ- 198
          || |:|| | ||||  |||||| :|||  ||||||||||||| |||:|||||:      |:
Db    65 LVSIQMWILLAGILVVYEAIDRIIAGPKNVDGFLMFLVSAFGLVVNIIMALLLGHDHGHRH 124

Qy   199 --------SGHGHSH---------------- 205
                  ||||||
Db   125 AGHSHGHGHGHGHSHGFTMSTHCDAKHTKDQHHHTHHHDENHPKDAHHTDEDHLHHA 184

Qy   206 ----------GTTNQQEENPSVRAAFIHVIGDFMQSMGVLVAAYILYFKPEYKYVD 251
                    |||  ||||||||||||||||||||||  |||: || ||||||||||
Db   185 HKEVTELLLGESKGGTKKKQWNINVQGAYLHVLGDSIQSIGVMIGGAVIWYNPRWQIVD 244

Qy   252 PICTFVFSILVLGTTLTLTILRDVILVLMEGTPKGVDFTAVRDLLLSVEGVEALHSLHIWAL 311
          ||||| ||:|||||||||| ||||::||:||||:|| |||:||||||| |||| ||| ||:|
Db   245 LICTLIFSVIVMGTTINMLRNIVLEVLMENTPREIDATKLERGLLDMEDVVAVHELHIWAI 304

Qy   312 TVAQPVLSVHIAIAQNTDAQAVLKTASSRLQGKFHFHTVTIQIE 355
          ||  |:|: |: |||||||||||||||||||| |||||||||||
Db   305 TVGKVLLACHVKIRREADADLVLDKVIDYIKRYNISHVTIOIE 348
```

```
Query Match         21.2%;
Best Local Similarity   28.4%;   Score 408.5;   DB 4;   Length 342;
Matches 97;   Conservative  80;   Mismatches 144;   Indels  21;   Gaps  7;

Qy    32 PLPRPGLDLQAIELAAQSNH-HCHAQKGPDSHCDPKKGKAQRQLYVASAICLLFMIGEVV 90
          |       |:||  |||  | |            |  | | | |||||| |||||||||||
Db    12 PFTTPNILVKEFLMTNINHQHHHSHAHGHVHHDNKK------VLMFSFTIISLFMLVEII 66

Qy    91 GGYLAHSLAVMTDAAHLLTDFASMLISLFSLWMSSRPATKTMNFGWQRAEILGALVSVLS 150
          ||:||||||| |||||||||||||:|||||||||||||||||||    |||||| || |
Db    67 GGFLANSLALLSDGFHMLSDAISLGVALIAFIYAEKHATKSKTYGVKRFEILAALFNGVT 126

Qy   151 IWVVTGVLVYLAVERLISGDYEIDGGTMLITSGCAVANIIMGLTLHQSGHGHSHGTTNQ 210
          |||| | | |||||||| |  ||:| || |||  || ||||| |  |||||||||    |
Db   127 LFIISIIIIIEAIRRFLEPP-EVQSKEMFIISVIGLMYNIIVAILMFKGG-------D 176

Qy   211 QEENPSVRAAFIHVIGDFMQSMGVLVAAYILYFKPEYKYVDPICTFVFSILVLGTTLTLTIL 270
          |||||||||||||||||||||| ||||||: || ||||||||||||||||||||||||||
Db   177 TSHNLMRGAFLHVLGDLFGSVGAIVAS-LLIWGFNFTIADPIASILVLSLIILKSAYGIS 235

Qy   271 RDVILVLMEGTPKGVDFTAVRDLLLSVEGVEALHSLHIWALTVAQPVLSVHIAIAQNTDA 330
          | |: ||  | | || |||:||||| || | | |   |:|||  |:|::|:||   :
Db   236 KSSINILMEGTPNDIDLNAVIKAISKDERIQNVHDCHVWTISNDMNALSCHAVVPEYLSV 295

Qy   331 Q----AVLKTASSRLQGKFHFHTVTIQIEDYSEDMKDCQACQG 369
            |     ||||||||| || |||||| |  :|  | ||||  |
Db   296 QTCETMLKSIESDLL-QLNIQIHMTIQLETPEHHDESTLCSG 336
```

RESULT 6
US-09-134-001C-5618
; Sequence 5618, Application US/09134001C
; Patent No. 6380370
; GENERAL INFORMATION:
; APPLICANT: Lynn Doucette-Stamm et al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCCUS
; TITLE OF INVENTION: EPIDERMIDIS FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: GTC-007
; CURRENT APPLICATION NUMBER: US/09/134,001C
; CURRENT FILING DATE: 1998-08-13
; PRIOR APPLICATION NUMBER: US 60/064,964
; PRIOR FILING DATE: 1997-11-08
; PRIOR APPLICATION NUMBER: US 60/055,779
; PRIOR FILING DATE: 1997-08-14
; NUMBER OF SEQ ID NOS: 5674
; SEQ ID NO 5618
; LENGTH: 359
; TYPE: PRT
; ORGANISM: Staphylococcus epidermidis

Query Match   20.0%; Score 385; DB 4; Length 359;
Best Local Similarity 27.7%; Pred. No. 1.1e-34;
Matches 88; Conservative 89; Mismatches 119; Indels 22; Gaps 7;

```
Qy   50 NHHCHA------QKGPD--SHCDPKGKAQRQ--LYVASACLLFMIGEVVGGYLAHSLAV 100
Db   37 NRRIHAWRRDMENKPDYFHHIEHRKFQSSKITLWLSLVITIMIFTVVEFVGGIVSNSLAL 96

Qy  101 MTDAAHLLTDFASMLISFSLWMSSRPATKTMNFGWQRAEILGALVSVLSIWVTGVLVY 160
Db   97 LSDSFHMLSDVLALGLSMVAIYFSSKPPTKNYTYGFLRLEIIVAFLNGLALIVISLGIMY 156

Qy  161 LAVERLISGDYEIDGGTMLITSGCAVAVNIIMGLTLHQSGHGHSHSGTTNQQEENPSVRAA 220
Db  157 EGIMRTIH-PRPVESGIMILIAFIGILIANIVLTILMIS-----LKKENNINIQSA 206

Qy  221 FIHVIGDFMQSMGVLVAAYILYFKPEYKYVDPICTFVFSILVLGTTLTLTILRDVILVLMEG 280
Db  207 LWHFIGDLLNSGIIVAFVLLHF-TGWNIVDPIISILISILILRGGYKIIKNASKVLMER 265

Qy  281 TPKGVDFTAVRDLLLSVEGVEALHSLHIWALTVAQPVLSVHIAIAQN--TDAQAVLKTAS 338
Db  266 VPDRDYDTEIMGAMKDVEGVIDIHEFHLWSVTTNQSSLSAHVVLSDDYIKSPYATINKVS 325

Qy  339 SRLQGKFHFHTYTIQIED 356
Db  326 DLLKTQYGLEHVTLQIEN 343
```

RESULT 7
US-09-461-474-14
; Sequence 14, Application US/09461474
; Patent No. 6278042
; GENERAL INFORMATION:
; APPLICANT: Allen, Steve
; APPLICANT: Rafalski, Antoni
; APPLICANT: Sakai, Hajime
; TITLE OF INVENTION: Plant Metal Transporters
; FILE REFERENCE: BB1303 US NA
; CURRENT APPLICATION NUMBER: US/09/461,474
; CURRENT FILING DATE: 1999-12-14
; EARLIER APPLICATION NUMBER: 60/112,562
; EARLIER FILING DATE: 1998-12-16
; NUMBER OF SEQ ID NOS: 17
; SOFTWARE: Microsoft Office 97
; SEQ ID NO 14
; LENGTH: 157
; TYPE: PRT
; ORGANISM: Triticum aestivum
; US-09-461-474-14

RESULT 8
US-09-134-001C-2867
; Sequence 2867, Application US/09134001C
; Patent No. 6380370
; GENERAL INFORMATION:
; APPLICANT: Lynn Doucette-Stamm et al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCC
; TITLE OF INVENTION: EPIDERMIDIS FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: GTC-007
; CURRENT APPLICATION NUMBER: US/09/134,001C
; CURRENT FILING DATE: 1998-08-13
; PRIOR APPLICATION NUMBER: US 60/064,964
; PRIOR FILING DATE: 1997-11-08
; PRIOR APPLICATION NUMBER: US 60/055,779
; PRIOR FILING DATE: 1997-08-14
; NUMBER OF SEQ ID NOS: 5674
; SEQ ID NO 2867
; LENGTH: 167
; TYPE: PRT
; ORGANISM: Staphylococcus epidermidis
; FEATURE:
; NAME/KEY: UNSURE
; LOCATION: (157)
; OTHER INFORMATION: Identity of amino acid at the above locations are unknown.
; US-09-134-001C-2867

Query Match   17.8%; Score 342; DB 4; Length 157;
Best Local Similarity 43.0%; Pred. No. 2e-30;
Matches 61; Conservative 40; Mismatches 41; Indels   0; Gaps 0;

```
Qy  214 NPSVRAAFIHVIGDFMQSMGVLVAAYILYFKPEYKYVDPICTFVFSILVLGTTLTLTILRDV 273
Db   14 NIVVHSAYLHVIGDSIQSIGVMIGGALIWYKPEWKIIDLICILIIFSVIVLFTTIRMIRNI 73

Qy  274 ILVLMEGTPKGVDFTAVRDLLLSVEGVEALHSLHIWALTVAQPVLSVHIAIAQNTDAQAV 333
Db   74 LEVLMESTPREIDATRLESGLRREMEGVIAVHELHIWAITVGKVLLACHVTTIQDADADKM 133

Qy  334 LKTASSRLQGKFHFHTYTIQIE 355
Db  134 LDKVIGYIKAEVNISHVTIQIE 155
```

RESULT 9
US-08-858-207A-520
; Sequence 520, Application US/08858207A
; Patent No. 6348328
; GENERAL INFORMATION:
; APPLICANT: Black, Michael
; APPLICANT: Hodgson, John
; APPLICANT: Knowles, David
; APPLICANT: Nicholas, Richard
; APPLICANT: Stodola, Robert
; TITLE OF INVENTION: Compounds
; NUMBER OF SEQUENCES: 552
; No. 6348328e1 Compounds
; CORRESPONDENCE ADDRESS:

Query Match   10.9%; Score 210.5; DB 4; Length 167;
Best Local Similarity 32.9%; Pred. No. 1.3e-15;
Matches 46; Conservative 30; Mismatches 63; Indels   1; Gaps 1;

```
Qy  216 SVRAAFIHVIGDFMQSMGVLVAAYILYFKPEYKYVDPICTFVFSILVLGTTLTLTILRDVIL 275
Db    4 STRRALVHVMGDLLGSVAGIIAGAVIYC-TGWMQIDPLLSVLVSLLLKSTFGVLRESYH 62

Qy  276 VLMEGTPKGVDFTAVRDLLLSVEGVEALHSLHIWALTVAQPVLSVHIAIAQNTDAQAVLK 335
Db   63 FLMEGVPLHIDIVVGDTQVDVDQVDGVIAVHDLHVWDMSPGQPALIGHVEIEHLDHWPNVLR 122

Qy  336 TASSRLQGKFHFHTYTIQIE 355
Db  123 AIKKMLLDKHGIDHITLQPE 142
```

```
;  ADDRESSEE: SmithKline Beecham Corporation
;  STREET: 709 Swedeland Road
;  CITY: King of Prussia
;  STATE: PA
;  COUNTRY: USA
;  ZIP: 19406-0939
;  COMPUTER READABLE FORM:
;  MEDIUM TYPE: Diskette
;  COMPUTER: IBM Compatible
;  OPERATING SYSTEM: DOS
;  SOFTWARE: FastSEQ for Windows Version 2.0
;  CURRENT APPLICATION DATA:
;  APPLICATION NUMBER: US/08/858,207A
;  FILING DATE: 09-MAY-1997
;  CLASSIFICATION: 435
;  PRIOR APPLICATION DATA:
;  APPLICATION NUMBER: 60/017670
;  FILING DATE: 14-MAY-1996
;  ATTORNEY/AGENT INFORMATION:
;  NAME: Gimmi, Edward R
;  REGISTRATION NUMBER: 38,891
;  REFERENCE/DOCKET NUMBER: P50475
;  TELECOMMUNICATION INFORMATION:
;  TELEPHONE: 610-270-4478
;  TELEFAX: 610-270-5090
;  TELEX:
;  INFORMATION FOR SEQ ID NO: 520:
;  SEQUENCE CHARACTERISTICS:
;  LENGTH: 165 amino acids
;  TYPE: amino acid
;  STRANDEDNESS: single
;  TOPOLOGY: linear
;  MOLECULE TYPE: No. 6348328e
US-08-858-207A-520

Query Match             6.7%;  Score 128;  DB 4;  Length 165;
Best Local Similarity  27.2%;  Pred. No. 2.3e-06;
Matches  47;  Conservative  39;  Mismatches  71;  Indels  16;  Gaps  5;

Qy   69 KAQRQLYVASAICLLFMIGEVVGGYLAHSLAVMTDAAHLLTDFASMLISLFSLWMSSRPA 128
        |:.:||  .||.|||.|.|.:||.|.||.:..|.|.|.|.||.||||.|.||.|.|||
Db    5 KAKYAVWWAFFLNLTYALVEFIAGGVFGCSSAVLADSVHDLGDAIAIGISAFLETISNREE 64

Qy  129 TKTMNFGWQRAEILGALVSVLSIWVVTG-VLVYLAVERLISGDYEIDGGTMLITSGCAVA 187
        |.:||  .::|||||.|...::||.|  .::.| |.|  .:|:.:|||||.|||.||||.||
Db   65 DNQYTLGYKRFSLLGALVT--AVILVTGSVLVILENVTKILHPQPVNDEGILWLGIIAIT 122

Qy  188 VNIMGLITLHQSGHGHSHGTTNQQEENPSVRAAFIHVIGDFMQSMGVLVAAYI 240
        :|.:||  .||    |.|    |  |    |:|   |  |.|  ::|:|.::|
Db  123 INLIASLVV---GKG------------KTKNESILS--LHFLEDTLGWVAVILMAIV 162
```

```
RESULT 10
US-09-134-001C-3700
;  Sequence 3700, Application US/09134001C
;  Patent No. 6380370
;  GENERAL INFORMATION:
;  APPLICANT: Lynn Doucette-Stamm et al
;  TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCCUS
;  TITLE OF INVENTION: EPIDERMIDIS FOR DIAGNOSTICS AND THERAPEUTICS
;  FILE REFERENCE: GTC-007
;  CURRENT APPLICATION NUMBER: US/09/134,001C
;  CURRENT FILING DATE: 1998-08-13
;  PRIOR APPLICATION NUMBER: US 60/064,964
;  PRIOR FILING DATE: 1997-11-08
;  PRIOR APPLICATION NUMBER: US 60/055,779
;  PRIOR FILING DATE: 1997-08-14
;  NUMBER OF SEQ ID NOS: 5674
;  SEQ ID NO 3700
;  LENGTH: 154
;  TYPE: PRT
;  ORGANISM: Staphylococcus epidermidis
;  FEATURE:
```

```
;  NAME/KEY: UNSURE
;  LOCATION: (109),(135),(143)
;  OTHER INFORMATION: Identity of amino acid at the above locations are unknown.
US-09-134-001C-3700

Query Match             5.2%;  Score 100.5;  DB 4;  Length 154;
Best Local Similarity  21.6%;  Pred. No. 0.0026;
Matches  27;  Conservative  35;  Mismatches  56;  Indels  7;  Gaps  1;

Qy   70 AQRQLYVASAICLLFMIGEVVGGYLAHSLAVMTDAAHLLTDFASMLISLFSLWMSSRPAT 129
        |:|| .|.||||.|.|.:..|..| .|.|..:|||||.|..:...:|.||||..||.||
Db   12 AQRGAFLSLVIYIILSIAKYMTGYVVDSAAVRADALNNMTDILVSVAVIVGLKISIKPAD 71

Qy  130 KTMNFGWQRAEILGALVSVLSIWVVTGVLVYLAVERLISGDYEIDGGTMLITSGCAVAVN 189
        .:.|  :|.:.:.|.:||:..  :::.:..:  |:..:|:..|..|.
Db   72 QNHPYGHLKSENISTL-----LVSFIIMFVGIQVIIENFPRLXTHDSHIPNFTIIIS 124

Qy  190 IIMGL 194
        .::|||
Db  125 MISGL 129
```

```
RESULT 11
US-08-405-271A-21
;  Sequence 21, Application US/08405271A
;  Patent No. 6432652
;  GENERAL INFORMATION:
;  APPLICANT: EVANS, CHRISTOPHER J.
;  APPLICANT: KEITH, DUANE E.
;  TITLE OF INVENTION: OPIOID RECEPTOR GENES
;  NUMBER OF SEQUENCES: 25
;  CORRESPONDENCE ADDRESS:
;  ADDRESSEE: MORRISON & FOERSTER
;  STREET: 2000 PENNSYLVANIA AVENUE, NW, Suite 5500
;  CITY: WASHINGTON
;  STATE: DC
;  COUNTRY: USA
;  ZIP: 20006-1888
;  COMPUTER READABLE FORM:
;  MEDIUM TYPE: Floppy disk
;  COMPUTER: IBM PC compatible
;  OPERATING SYSTEM: PC-DOS/MS-DOS
;  SOFTWARE: Patentin Release #1.0, Version #1.30
;  CURRENT APPLICATION DATA:
;  APPLICATION NUMBER: US/08/405,271A
;  FILING DATE: 14-MAR-1995
;  CLASSIFICATION: 435
;  ATTORNEY/AGENT INFORMATION:
;  NAME: MURASHIGE, KATE H.
;  REGISTRATION NUMBER: 29,959
;  REFERENCE/DOCKET NUMBER: 22000-20526.22
;  TELECOMMUNICATION INFORMATION:
;  TELEPHONE: (202) 887-1500
;  TELEFAX: (202) 887-0763
;  TELEX: 90-4030 MRSNFOERSWSH
;  INFORMATION FOR SEQ ID NO: 21:
;  SEQUENCE CHARACTERISTICS:
;  LENGTH: 370 amino acids
;  TYPE: amino acid
;  STRANDEDNESS: single
;  TOPOLOGY: linear
;  FEATURE:
;  NAME/KEY: Modified-site
;  LOCATION: group(18, 33)
;  OTHER INFORMATION: /note= "extracellular Asn residues
;  OTHER INFORMATION: that are consensus sites for N-linked glycosylation"
US-08-405-271A-21

Query Match             5.2%;  Score 100.5;  DB 4;  Length 370;
Best Local Similarity  21.2%;  Pred. No. 0.01;
Matches  83;  Conservative  59;  Mismatches  137;  Indels  113;  Gaps  20;

Qy   32 PLPRPGLDLQAIELAAQSNHHCHA--QKGPDSHCDPKKGKAQRQLYVA-----SAICLLF 84
```

```
Db     3  PAPSAGAELQPPLFANASDAYPSACPSAGANASGPPARSASSLALAIATALYSAVCAG
          ||  :|:|                                     |||:||;
Qy    85  MIGEVGGYLAHSLAVMTDAAHLITDFASMLISLFSLWMSSRPAITKTMNF-------GW 136
          ||:|| |: | :||  :||  :|| ::|
Db    63  LIGNVL---------VMFGIVRYRTMKTATNIYIFNLALADALATSTLPFQSAKYLMETW 113

Qy   137  QRAELL-GALVSV------LSIWVTGVLV--YLAVERLISG-DYEIDGGTMLI------ 180
          ::||:  |:||           ||  |||  ||   ||
Db   114  PFGELLCKAVLSIDYYNMFTSIFTLTMMSVDRYIAVCHPVKALDFRTPAKALINICIWV 173

Qy   181  -TSGCAVANIIMGLTLHQSGHGHSHGTINQQEENPSVRAAFIHVIGDFMQSMGVLVAAY 239
           :|  |  |  |   |                               :| |
Db   174  LASGVGVPI-MVMAVTRPRDG---------------------------------AVVCM 198

Qy   240  ILYFKPEYKY---VDPICTFVFSILVLVGTTLTILRDVILV-----LMEGT-PKGVDFTAV 290
          |   :     |  ||:  |||| ::|                  |:
Db   199  LQFPSPSWWDTVTKICVFLFAFVVPILITVCYGIMLLRLRSVRLLSGSKEKDRSLRRI 258

Qy   291  RDLLLSVEG--VEALHSLHI----WALTVAQ---PVLSVHIAIAQNTDAQAVLKTASSR 340
          : :|  |     ||| |         || |         |:|    |
Db   259  TRMVLVVVGAFVVCWNAPIHIFVIVWTLVDIRRDPLVVAALHLCIA-------LGYANSS 310

Qy   341  LQGKFHFHTVTIQIEDYSEDMKDC---QACQGP 370
          | |
Db   311  LN-------PVLYAFLDENFKRCFRQLCRKP 334
```

RESULT 12
US-09-134-001C-5241
; Sequence 5241, Application US/09134001C
; Patent No. 6380370
; GENERAL INFORMATION:
; APPLICANT: Lynn Doucette-Stamm et al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCCUS
; TITLE OF INVENTION: EPIDERMIDIS FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: GTC-007
; CURRENT APPLICATION NUMBER: US/09/134,001C
; CURRENT FILING DATE:
; PRIOR APPLICATION NUMBER: 1998-08-13
; PRIOR FILING DATE: 1997-11-08
; PRIOR APPLICATION NUMBER: US 60/064,964
; PRIOR FILING DATE: 1997-08-14
; PRIOR APPLICATION NUMBER: US 60/055,779
; NUMBER OF SEQ ID NOS: 5674
; SEQ ID NO 5241
; LENGTH: 489
; TYPE: PRT
; ORGANISM: Staphylococcus epidermidis
US-09-134-001C-5241

Query Match   4.6%; Score 88; DB 4; Length 489;
Best Local Similarity 19.9%; Pred. No. 0.4;
Matches 61; Conservative 53; Mismatches 119; Indels 74; Gaps 12;

```
Qy    67  KGKAQRQLYVASAICL----LFMIGEVGGYLAHSLAVMTDAAHLITDFASMLISLFSLW 122
          |||                  ||:|| |: | :||  :||  :|| ::|
Db   153  KGLNQKWLGIVFAVLITVTFAFVFNTVQANTIAESLNTQYNISPVITGIVLAVI----- 206

Qy   123  MSSRPAITKTMNFGWQR--AEILGALVSVLSIWVVTGVLVYLA--VERLIS-------- 168
          |:||  :                                       ::|:
Db   207  -----TGIIFPGGVRSIATLSSLIVPIMAIVYIGMVLILLLNIDQIVPMIGTIIKSAF 260

Qy   169  GDYEIDGGTMLITSGCAVANIIMGLTLHQSGHGHSHGTINQQEENPSVRAAFIHVIGD- 227
                      :| | | |
Db   261  GVQQVTGGAV---GAAILQGIKRGLFSNEAGMGSA-------PNAAATSAVPHPVKQG 308

Qy   228  FMQSMGVLVAAYILYFKPEYKYVDPICTFVFSILVLVGTTLTILRDVILVLMEGTPKGVDF 287
                | |
Db   309  LIQSLGVFFDTML---------VCT------ATAIMILLYSGLQFGDSAPQGVAV 348

Qy   288  T--AVRDLILLSVEGVEALHSLHIWALTVAQPVLSVHIAIAQNTDAQAVLKTASSRLQGKF 345
          |   :: | |
Db   349  TQSALNEHLGSAGGI-------FLTVAVLFAFSSVVGNYYGQSNIEFLSNNKMILF 399
```

```
Qy   346  HFHTVTI 352
          :
Db   400  IFRCFVV 406
```

RESULT 13
US-09-134-001C-3689
; Sequence 3689, Application US/09134001C
; Patent No. 6380370
; GENERAL INFORMATION:
; APPLICANT: Lynn Doucette-Stamm et al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCC
; TITLE OF INVENTION: EPIDERMIDIS FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: GTC-007
; CURRENT APPLICATION NUMBER: US/09/134,001C
; CURRENT FILING DATE: 1998-08-13
; PRIOR APPLICATION NUMBER: US 60/064,964
; PRIOR FILING DATE: 1997-11-08
; PRIOR APPLICATION NUMBER: US 60/055,779
; PRIOR FILING DATE: 1997-08-14
; NUMBER OF SEQ ID NOS: 5674
; SEQ ID NO 3689
; LENGTH: 766
; TYPE: PRT
; ORGANISM: Staphylococcus epidermidis
US-09-134-001C-3689

Query Match   4.5%; Score 86.5; DB 4; Length 766;
Best Local Similarity 21.2%; Pred. No. 1.2;
Matches 61; Conservative 50; Mismatches 94; Indels 83; Gaps 12;

```
Qy    88  EVVGGY--------LAHSLAVMTDAAHLLTD----------------------FASML-IS 117
          |  ||         ||:||||:||  |                          ||:| :|
Db   219  EISGGFNGKKGVEEAKQIAELLNAGSLPVDLKEIYSNSVGAQFGQDALDKTMFASIVGIA 278

Qy   118  LFSLWMSSRPAITKTMNFGWQRAEILGALVSVLSIWVVTGVLVYLAVERLISGDYEIDG-G 176
          |   :                 : |:|| |  :    |||: |  |  |
Db   279  LIYLFM----------LGFYR--LPGLVAIAL--TTYIYLTLVAFNFISGVTLPGLA 323

Qy   177  TMLITSGCAVANIIMGLTLHQSGHGHSHGTINQQEENPSVRAAFIHVIGDFMQSMGVLV 236
          :|: |  |  | ||    |                               |
Db   324  ALVLGVGMAVDANIIMYERIKDE--LRIGRILKQAYSKANKSSFLII--FDSNLITVI 377

Qy   237  AAYILYFKPEYKYVDPICTFVFSILVLVGTTLTILRDVILVLMEGT------------ 281
          ||  |:
Db   378  AAAVLFFFGESSVKGFATMLLLGIIMIFVTAVFLSRGLLSLLVSSNFPKKQYWLFGVKKM 437

Qy   282  ----PKGVDFTAVRDLILLSVEGVEALHSLHIWALTVAQPVLSVHIAI 324
              |:  |:    |:  |:     :||| ||:  | |
Db   438  DRHDINEGKD--VHDLKTSYERLN-----FVKLAKPLISLSILI 474
```

RESULT 14
US-09-268-866-2
; Sequence 2, Application US/09268866
; Patent No. 6316272
; GENERAL INFORMATION:
; APPLICANT: Mack, David
; APPLICANT: Gish, Kurt C.
; APPLICANT: Wilson, Keith E.
; TITLE OF INVENTION: No. 6316272el Methods of Diagnosing of Colorectal Cancer,
; TITLE OF INVENTION: Compositions, and Methods of Screening for Colorectal
; TITLE OF INVENTION: Cancer Modulators
; FILE REFERENCE: A67474/RMS/DAV/JJD
; CURRENT APPLICATION NUMBER: US/09/268,866
; CURRENT FILING DATE: 1999-03-15
; NUMBER OF SEQ ID NOS: 2
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 2
; LENGTH: 1212
; TYPE: PRT
; ORGANISM: Homo sapiens

```
US-09-268-866-2

Query Match    4.48;   Score 85.5;  DB 4;   Length 1212;
Best Local Similarity  22.6%;  Pred. No. 3.2;
Matches  48;  Conservative  45;  Mismatches  60;  Indels  59;  Gaps  14;

Qy   57 KGPDSHCDPKRGKAQRQLYVASAICLLFMIGEVVGGYLAHSLAVMTDAAHLLTDF--ASM 114
          |:|  :|| :: | |||:::|  ||:|:|||| |  | |:|:::|| |||||||:  |
Db  643 KGYGKNNEPLRG------------YI----LTFLI-ALGFILIAELNVI--APIISNFFLASY 686

Qy  115 LISLFSLWMSS-------RPATKTMNFGWQRAEILGALVS-----VLSIW-----VVT 155
         | |:|:|:   ::|     :|||:  |:|  |||  : ||        |
Db  687 ALINFSVFHASLAKSPGWRPAFKYYNM-W--ISLLGAILCCIVMFVINWWAALLTYIVL 743

Qy  156 GVLVYLAVERLISGDYEIDGGTMLITSGCAVANI-IMGLTLHQSGHGHSHGTTNQQEE- 213
         | |: : :  |  :|:| | ||||||:|||||| || |||:||:|||:|   ::|
Db  744 PDVNWGSSTQALTYLNALQHSIRLSGVEDH------------VKNFRPQC 791

Qy  214 -----NPSVRAAFIHVIGDFMQSMGVLVAAYI 240
          |  |:||:|| | ||:|| | ::|:|:|||:
Db  792 LVMTGAPNSRPALLHLVHDFTRNVGLMICGHV 823
```

RESULT 15
US-09-335-409-5
Sequence 5, Application US/09335409
Patent No. 612029
GENERAL INFORMATION:
  APPLICANT: Schupp, Thomas
  APPLICANT: Ligon, James
  APPLICANT: Molnar, Istvan
  APPLICANT: Zirkle, Ross
  APPLICANT: Cyr, Devon
  APPLICANT: Goerlach, Joern
  TITLE OF INVENTION: GENES FOR THE BIOSYNTHESIS OF EPOTHILONES
  FILE REFERENCE: 4-30582A
  CURRENT APPLICATION NUMBER: US/09,335,409
  CURRENT FILING DATE: 1999-06-17
  NUMBER OF SEQ ID NOS: 30
  SOFTWARE: Patentin Ver. 2.0
  SEQ ID NO 5
  LENGTH: 7257
  TYPE: PRT
  ORGANISM: Sorangium cellulosum
US-09-335-409-5

```
Query Match    4.48;   Score 85.5;  DB 3;   Length 7257;
Best Local Similarity  18.6%;  Pred. No. 54;
Matches  84;  Conservative  46;  Mismatches  126;  Indels  195;  Gaps  19;

Qy   24 WQEGAGWIPLPRPGLDLQAIELAAQSNHHCHAQKGPDSHCDPKKGKAQRQLYVASAICLL 83
          || |  |   | | ||  :| ::  |:| ||   :| | |:|:||:||||:|:|  |:|
Db  902 WQRQRYWIEAPAEGLGATAADALAQWFYRVDWPEMPRSSVDSRRARSGGWLVLADR---- 957

Qy   84 FMIGEVVGGYLA------HSLAVMTDAAHLLT------DFASMLISLFSLW---- 122
          | || | : |       |||||||| |||||      ::|:|||| :|:|
Db  958 GGVGEAAAAALSSQGCSCAVLHAPAEASAVAEQVTQALGGRNDWQGVLY----LWGLDAV 1013

Qy  123 ---MSSRPATKTMNFGWQRAEILGALVSVLS-----IWVVT------ 155
          :|::|| |   ||| |:| ||:| | ::        ||||:
Db 1014 VEAGASAEEVAKVTHLA--AAPVL-ALIQALGTGPRSPRLMIVTRGACTVGGEPDAAPCQ 1070

Qy  156 -------GVLVYL-----AVERLISGDYEID----- 174
                |:|:|:     |||||||||:|
Db 1071 AALWGMGRVAALEHPGSWGGLVDLDPEESPTEVEALVAELLSPDAEDQLAFRQGRRRAAR 1130

Qy  175 -------GGTMLITSGCAVANVNIIMGLTLHQSGHGH----- 212
                |||||||||||| :|| :||:||||:|||
Db 1131 LVAAPPEGNAAPVSLSAEGSYILVTGGLG-ALGLLVARWLVERGAGHLVLISRHGLPDREE 1189

Qy  213 ---ENPSVRAAFIHVIGDFMQSMGVLVAAYILYFKPEYKYVDPICTFVFSILVLGTTLT 268
          :  :||:||:| |||:|||| ::|:|:|||
Db 1190 WGRDQPPEVRARIAIAIEALEAQGARVTVAA------- 1219
```

RESULT 16
US-09-568-102-5
Sequence 5, Application US/09568102
Patent No. 6346404
GENERAL INFORMATION:
  APPLICANT: Schupp, Thomas
  APPLICANT: Ligon, James
  APPLICANT: Molnar, Istvan
  APPLICANT: Zirkle, Ross
  APPLICANT: Cyr, Devon
  APPLICANT: Goerlach, Joern
  TITLE OF INVENTION: GENES FOR THE BIOSYNTHESIS OF EPOTHILONES
  FILE REFERENCE: 4-30582A
  CURRENT APPLICATION NUMBER: US/09/568,102
  CURRENT FILING DATE: 2000-05-10
  PRIOR APPLICATION NUMBER: 09/335,409
  PRIOR FILING DATE: 1999-06-17
  NUMBER OF SEQ ID NOS: 30
  SOFTWARE: Patentin Ver. 2.0
  SEQ ID NO 5
  LENGTH: 7257
  TYPE: PRT
  ORGANISM: Sorangium cellulosum
US-09-568-102-5

```
Query Match    4.48;   Score 85.5;  DB 4;   Length 7257;
Best Local Similarity  18.6%;  Pred. No. 54;
Matches  84;  Conservative  46;  Mismatches  126;  Indels  195;  Gaps  19;

Qy   24 WQEGAGWIPLPRPGLDLQAIELAAQSNHHCHAQKGPDSHCDPKKGKAQRQLYVASAICLL 83
          || |  |   | | ||  :| ::  |:| ||   :| | |:|:||:||||:|:|  |:|
Db  902 WQRQRYWIEAPAEGLGATAADALAQWFYRVDWPEMPRSSVDSRRARSGGWLVLADR---- 957

Qy   84 FMIGEVVGGYLA------HSLAVMTDAAHLLT------DFASMLISLFSLW---- 122
          | || | : |       |||||||| |||||      ::|:|||| :|:|
Db  958 GGVGEAAAAALSSQGCSCAVLHAPAEASAVAEQVTQALGGRNDWQGVLY----LWGLDAV 1013

Qy  123 ---MSSRPATKTMNFGWQRAEILGALVSVLS-----IWVVT------ 155
          :|::|| |   ||| |:| ||:| | ::        ||||:
Db 1014 VEAGASAEEVAKVTHLA--AAPVL-ALIQALGTGPRSPRLMIVTRGACTVGGEPDAAPCQ 1070

Qy  156 -------GVLVYL-----AVERLISGDYEID----- 174
                |:|:|:     |||||||||:|
Db 1071 AALWGMGRVAALEHPGSWGGLVDLDPEESPTEVEALVAELLSPDAEDQLAFRQGRRRAAR 1130

Qy  175 -------GGTMLITSGCAVANVNIIMGLTLHQSGHGH----- 212
                |||||||||||| :|| :||:||||:|||
Db 1131 LVAAPPEGNAAPVSLSAEGSYILVTGGLG-ALGLLVARWLVERGAGHLVLISRHGLPDREE 1189

Qy  213 ---ENPSVRAAFIHVIGDFMQSMGVLVAAYILYFKPEYKYVDPICTFVFSILVLGTTLT 268
          :  :||:||:| |||:|||| ::|:|:|||
Db 1190 WGRDQPPEVRARIAIAIEALEAQGARVTVAA------- 1219

Qy  269 ILRDVILVLMEGTPKGVDFTAVRDLLLSVEGVEALHSLHIWALTVAQPVLS--VHIA--- 323
          | :  :|:| |||:|||||||| :|  |:  |:   |||||||: || |
Db 1220 ------------VD------VADAEGMAAL------LAAVEPPLRGVVHAAGLL 1249

Qy  324 ---IAQNTDAQAVLKTASSRLQGKFHFHTVT 351
          :  |||:|||||:|:|:|::||:|  | |||
Db 1250 DDGLLAHQDAGRLARVLRPKVEGAWVLHTLT 1280
```

RESULT 17

US-09-567-969-5
Sequence 5, Application US/09567969
; Patent No. 6355457
; GENERAL INFORMATION:
; APPLICANT: Schupp, Thomas
; APPLICANT: Ligon, James
; APPLICANT: Molnar, Istvan
; APPLICANT: Zirkle, Ross
; APPLICANT: Cyr, Devon
; APPLICANT: Goerlach, Joern
; TITLE OF INVENTION: GENES FOR THE BIOSYNTHESIS OF EPOTHILONES
; FILE REFERENCE: 4-30582A
; CURRENT APPLICATION NUMBER: US/09/567,969
; CURRENT FILING DATE: 2000-05-10
; PRIOR APPLICATION NUMBER: 09/335,409
; PRIOR FILING DATE: 1999-06-17
; NUMBER OF SEQ ID NOS: 30
; SOFTWARE: PatentIn Ver. 2.0
;
; SEQ ID NO 5
; LENGTH: 7257
; TYPE: PRT
; ORGANISM: Sorangium cellulosum
US-09-567-969-5

Query Match   4.4%; Score 85.5; DB 4; Length 7257;
Best Local Similarity   18.6%; Pred. No. 54;
Matches   84; Conservative   46; Mismatches 126; Indels 195; Gaps 19;

Qy   24 WQEGAGWIPLPRPGLDLQAIELAAQSNHHCHAQKGPDSHCDPKKGKAQRQLYVASAICLL 83
              || |||| || ||| | :|| | ||| :| ::| |: : || | || |:|||
Db   902 WQRQRYWIEAPAEGLGATAADALAQWFYRVDWPEMPRSSVDSRRARSGGWLVLADR---- 957

Qy   84 FMIGEVVGGYLA------- 122
              ||| || |:
Db   958 GGVGEAAAAALSSQGCSCAVLHAPAEASAVAEQVTQALGGRNDWQGVLY----LWGLDAV 1013

Qy   123 ---MSSRPATKTMNFGWQRAEILGALVSVLS-------IWVVT------ 155
              |:| | | |:|| | : ||| |:   :| ||
Db   958 GGVGEAAAAALSSQGCSCAVLHAPAEASAVAEQVTQALGGRNDWQGVLY----LWGLDAV 1013

Qy   84 FMIGEVVGGYLA------- 122
              ||| || |:
Db   958 HSLAVMTDAAHLLT-------DFASMLISLFSLW----- 122

Qy   123 ---MSSRPATKTMNFGWQRAEILGALVSVLS-------IWVVT------ 155
              |:| | | |:|| | : ||| |:   :| ||
Db   1014 VEAGASAEEVAKVTHLA--AAPVL-ALIQALGTGPRSPRLWIVTRGACTVGGGEPDAAPCQ 1070

Qy   156 -------GVLYL----------AVERLISGDYEID--------- 174
                      |:| || ::| | |
Db   1071 AALWGMGRVAALEHPGSWGGLVDLDPEESPTEVEALVAELLSPDAEDQLAFRQGRRRAAR 1130

Qy   175 -------GGTMLITSGCAVAVNIIMGLTLHQSGHGH-----SHGTTNQQE 212
                      | |:| || || : || || | ::|
Db   1131 LVAAPPEGNAAPVSLSAEGSYLVTGGLG-ALGLLVARWLVERGAGHLVLISRHGLPDREE 1189

Qy   213 -----ENPSVRAAFIHVIGDFMQSMGVLVAAYILYFKPEYKYVDPICTFVFSILVLGTTLT 268
                   || | :| ||| | ||| | | |:|:| | | | | ||:|| |
Db   1190 WGRDQPPEVRARIAAIEALEAQGARVTVAA------------------------- 1219

Qy   269 ILRDVILVLMEGTPKGVDFTAVRDLLLSVEGVEALHSLHIWALTVAQPVLS--VHIA--- 323
              :|::   :| ||| ||| ||: |||| || | : | | | |:| :
Db   1220 ---------VD------VADAEGMAAL-------LAAVEPPLRGVVHAAGLL 1249

Qy   324 ---IAQNTDAQAVLKTASSRLQGKFHFHTVT 351
                   | | |:|||:|: || | :| ::|
Db   1250 DDGLLAHQDAGRLIARVLRPKVEGAWVLHTLT 1280

RESULT 18
US-09-568-480-5
Sequence 5, Application US/09568480
; Patent No. 6355458
; GENERAL INFORMATION:
; APPLICANT: Schupp, Thomas
; APPLICANT: Ligon, James
; APPLICANT: Molnar, Istvan
; APPLICANT: Zirkle, Ross
; APPLICANT: Cyr, Devon
; APPLICANT: Goerlach, Joern
; TITLE OF INVENTION: GENES FOR THE BIOSYNTHESIS OF EPOTHILONES

; FILE REFERENCE: 4-30582A
; CURRENT APPLICATION NUMBER: US/09/568,480
; CURRENT FILING DATE: 2000-05-10
; PRIOR APPLICATION NUMBER: 09/335,409
; PRIOR FILING DATE: 1999-06-17
; NUMBER OF SEQ ID NOS: 30
; SOFTWARE: PatentIn Ver. 2.0
;
; SEQ ID NO 5
; LENGTH: 7257
; TYPE: PRT
; ORGANISM: Sorangium cellulosum
US-09-568-480-5

Query Match   4.4%; Score 85.5; DB 4; Length 7257;
Best Local Similarity   18.6%; Pred. No. 54;
Matches   84; Conservative   46; Mismatches 126; Indels 195; Gaps 19;

Qy   24 WQEGAGWIPLPLPRPGLDLQAIELAAQSNHHCHAQKGPDSHCDPKKGKAQRQLYVASAICLL 83
              || |||| || ||| | :|| | ||| :| ::| |: : || | || |:|||
Db   902 WQRQRYWIEAPAEGLGATAADALAQWFYRVDWPEMPRSSVDSRRARSGGWLVLADR---- 957

Qy   84 FMIGEVVGGYLA------- 122
              ||| || |:
Db   958 GGVGEAAAAALSSQGCSCAVLHAPAEASAVAEQVTQALGGRNDWQGVLY----LWGLDAV 1013

Qy   84 ----------HSLAVMTDAAHLLT-------DFASMLISLFSLW----- 122
                        | | | :|:|| | : || |:
Db   1014 VEAGASAEEVAKVTHLA--AAPVL-ALIQALGTGPRSPRLWIVTRGACTVGGGEPDAAPCQ 1070

Qy   123 ---MSSRPATKTMNFGWQRAEILGALVSVLS-------IWVVT------ 155
              |:| | | |:|| | : ||| |:   :| ||
Db   958 GGVGEAAAAALSSQGCSCAVLHAPAEASAVAEQVTQALGGRNDWQGVLY----LWGLDAV 1013

Qy   156 -------GVLYL----------AVERLISGDYEID--------- 174
                      |:| || ::| | |
Db   1071 AALWGMGRVAALEHPGSWGGLVDLDPEESPTEVEALVAELLSPDAEDQLAFRQGRRRAAR 1130

Qy   175 -------GGTMLITSGCAVAVNIIMGLTLHQSGHGH-----SHGTTNQQE 212
                      | |:| || || : || || | ::|
Db   1131 LVAAPPEGNAAPVSLSAEGSYLVTGGLG-ALGLLVARWLVERGAGHLVLISRHGLPDREE 1189

Qy   213 -----ENPSVRAAFIHVIGDFMQSMGVLVAAYILYFKPEYKYVDPICTFVFSILVLGTTLT 268
                   || | :| ||| | ||| | | |:|:| | | | | ||:|| |
Db   1190 WGRDQPPEVRARIAAIEALEAQGARVTVAA------------------------- 1219

Qy   269 ILRDVILVLMEGTPKGVDFTAVRDLLLSVEGVEALHSLHIWALTVAQPVLS--VHIA--- 323
              :|::   :| ||| ||| ||: |||| || | : | | | |:| :
Db   1220 ---------VD------VADAEGMAAL-------LAAVEPPLRGVVHAAGLL 1249

Qy   324 ---IAQNTDAQAVLKTASSRLQGKFHFHTVT 351
                   | | |:|||:|: || | :| ::|
Db   1250 DDGLLAHQDAGRLIARVLRPKVEGAWVLHTLT 1280

RESULT 19
US-09-568-486-5
Sequence 5, Application US/09568486
; Patent No. 6355459
; GENERAL INFORMATION:
; APPLICANT: Schupp, Thomas
; APPLICANT: Ligon, James
; APPLICANT: Molnar, Istvan
; APPLICANT: Zirkle, Ross
; APPLICANT: Cyr, Devon
; APPLICANT: Goerlach, Joern
; TITLE OF INVENTION: GENES FOR THE BIOSYNTHESIS OF EPOTHILONES
; FILE REFERENCE: 4-30582A
; CURRENT APPLICATION NUMBER: US/09/568,486
; CURRENT FILING DATE: 2000-05-10
; PRIOR APPLICATION NUMBER: 09/335,409
; PRIOR FILING DATE: 1999-06-17
; NUMBER OF SEQ ID NOS: 30
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 5
; LENGTH: 7257
; TYPE: PRT
; ORGANISM: Sorangium cellulosum

US-09-568-486-5

```
Query Match       4.4%;     Score 85.5;   DB 4;   Length 7257;
Best Local Similarity   18.6%;   Pred. No. 54;
Matches   84;   Conservative   46;   Mismatches   126;   Indels   195;   Gaps   19;

Qy     24  WQEGAGWIPLPRPGLDLQAIELAAQSNHHCHAQKGPDSHCDPKKGKAQRQLYVASAICLL    83
               |||  ||:||  |:|  :|:||        |:  |||||:|||
Db    902  WQRQRYWIEAPAEGLGATAADALAQWFYRVDWPEMPRSSVDSRRARSGGWLIVLADR----   957

Qy     84  FMIGEVVGGYLA--------HSLAVMTDAAHLLT------DFASMLISLFSLW-----    122
               :|||              |  :|                |||:|
Db    958  GGVGEAAAAALSSQGCSCAVLHAPAEASAVAEQVTQALGGRNDWQGVLY----LWGLDAV   1013

Qy    123  ------MSSRPATKTMNFGWQRAEILGALVSVLS------IWVVT----------    155
               |:|                                  |:|:
Db   1014  VEAGASAEEVAKVTHLA--AAPVL-ALIQALGTGPRSPRLWIVTRGACTVGGEPDAAPCQ   1070

Qy    156  -------------GVLVYL-----------AVERLISGDYEID------------    174
                         | ||             |:| |
Db   1071  AALWGMGRVAALEHPGSWGGLVDLDPEESPTEVEALVAELLSPDAEDQLAFROGRRRAAR   1130

Qy    175  ------------GGTMLITSGCAVANVNIIMGLTLHQSGHGH-----SHGTTNQOE    212
                       |:|:|                              ||  :||
Db   1131  LVAAPPEGNAAPVSLSAEGSYLVTGGLG-ALGLLVARWLVERGAGHLVLISRHGLPDREE   1189

Qy    213  ----ENPSVRAAFIHVICDFMQSMGVLVAAYILYFKPEYKYVDPICTFVFSILVLGTTLT   268
               :|:                |:||:
Db   1190  WGRDQPPEVRARIAAIEALEAQGARVTVAA--------------------------   1219

Qy    269  ILRDVILVLMEGTPKGVDFTAVRDLLLSVEGVEALHSLHIWALTIVAQPVLS--VHIA---   323
               ||      |:||                        ||:|          |||:
Db   1220  ----VD------VADAEGMAAL-----LAAVEPPLRGVVHAAGLL   1249

Qy    324  ---IAQNTDAQAVLKTASSRLQGKFHFHTVT   351
              :  ||     |||| |:|:
Db   1250  DDGLLAHODAGRLARVLRPKVEGAWLHTLT   1280
```

RESULT 20
US-09-568-472-5
Sequence 5, Application US/09568472
; Patent No. 6358719
; GENERAL INFORMATION:
;   APPLICANT: Schupp, Thomas
;   APPLICANT: Ligon, James
;   APPLICANT: Molnar, Istvan
;   APPLICANT: Zirkle, Ross
;   APPLICANT: Cyr, Devon
;   APPLICANT: Goerlach, Joern
;   TITLE OF INVENTION: GENES FOR THE BIOSYNTHESIS OF EPOTHILONES
;   FILE REFERENCE: 4-30582A
;   CURRENT APPLICATION NUMBER: US/09/568,472
;   CURRENT FILING DATE: 2000-05-10
;   PRIOR APPLICATION NUMBER: 09/335,409
;   PRIOR FILING DATE: 1999-06-17
;   NUMBER OF SEQ ID NOS: 30
;   SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 5
;   LENGTH: 7257
;   TYPE: PRT
;   ORGANISM: Sorangium cellulosum
US-09-568-472-5

```
Query Match       4.4%;     Score 85.5;   DB 4;   Length 7257;
Best Local Similarity   18.6%;   Pred. No. 54;
Matches   84;   Conservative   46;   Mismatches   126;   Indels   195;   Gaps   19;

Qy     24  WQEGAGWIPLPRPGLDLQAIELAAQSNHHCHAQKGPDSHCDPKKGKAQRQLYVASAICLL    83
               |||  ||:||  |:|  :|:||        |:  |||||:|||
Db    902  WQRQRYWIEAPAEGLGATAADALAQWFYRVDWPEMPRSSVDSRRARSGGWLIVLADR----   957

Qy     84  FMIGEVVGGYLA--------HSLAVMTDAAHLLT------DFASMLISLFSLW-----    122
```

RESULT 21
US-09-567-899-5
; Sequence 5, Application US/09567899
; Patent No. 6383787
; GENERAL INFORMATION:
;   APPLICANT: Schupp, Thomas
;   APPLICANT: Ligon, James
;   APPLICANT: Molnar, Istvan
;   APPLICANT: Zirkle, Ross
;   APPLICANT: Cyr, Devon
;   APPLICANT: Goerlach, Joern
;   TITLE OF INVENTION: GENES FOR THE BIOSYNTHESIS OF EPOTHILONES
;   FILE REFERENCE: 4-30582A
;   CURRENT APPLICATION NUMBER: US/09/567,899
;   CURRENT FILING DATE: 2000-05-10
;   PRIOR APPLICATION NUMBER: 09/335,409
;   PRIOR FILING DATE: 1999-06-17
;   NUMBER OF SEQ ID NOS: 30
;   SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 5
;   LENGTH: 7257
;   TYPE: PRT
;   ORGANISM: Sorangium cellulosum
US-09-567-899-5

```
Query Match       4.4%;     Score 85.5;   DB 4;   Length 7257;
Best Local Similarity   18.6%;   Pred. No. 54;
Matches   84;   Conservative   46;   Mismatches   126;   Indels   195;   Gaps   19;

Qy     24  WQEGAGWIPLPRPGLDLQAIELAAQSNHHCHAQKGPDSHCDPKKGKAQRQLYVASAICLL    83
               |||  ||:||  |:|  :|:||        |:  |||||:|||
Db    902  WQRQRYWIEAPAEGLGATAADALAQWFYRVDWPEMPRSSVDSRRARSGGWLIVLADR----   957

Qy     84  FMIGEVVGGYLA--------HSLAVMTDAAHLLT------DFASMLISLFSLW-----    122
               :|||              |  :|                |||:|
Db    958  GGVGEAAAAALSSQGCSCAVLHAPAEASAVAEQVTQALGGRNDWQGVLY----LWGLDAV   1013

Qy    123  ------MSSRPATKTMNFGWQRAEILGALVSVLS------IWVVT----------    155
               |:|                                  |:|:
Db   1014  VEAGASAEEVAKVTHLA--AAPVL-ALIQALGTGPRSPRLWIVTRGACTVGGEPDAAPCQ   1070

Qy    156  -------------GVLVYL-----------AVERLISGDYEID------------    174
                         | ||             |:| |
Db   1071  AALWGMGRVAALEHPGSWGGLVDLDPEESPTEVEALVAELLSPDAEDQLAFROGRRRAAR   1130

Qy    175  ------------GGTMLITSGCAVANVNIIMGLTLHQSGHGH-----SHGTTNQOE    212
                       |:|:|                              ||  :||
Db   1131  LVAAPPEGNAAPVSLSAEGSYLVTGGLG-ALGLLVARWLVERGAGHLVLISRHGLPDREE   1189

Qy    213  ----ENPSVRAAFIHVIGDFMQSMGVLVAAYILYFKPEYKYVDPICTFVFSILVLGTTLT   268
               :|:                |:||:
Db   1190  WGRDQPPEVRARIAAIEALEAQGARVTVAA---   1219

Qy    269  ILRDVILVLMEGTPKGVDFTAVRDLLLSVEGVEALHSLHIWALTIVAQPVLS--VHIA---   323
               ||      |:||                        ||:|          |||:
Db   1220  ----VD------VADAEGMAAL-----LAAVEPPLRGVVHAAGLL   1249

Qy    324  ---IAQNTDAQAVLKTASSRLQGKFHFHTVT   351
              :  ||     |||| |:|:
Db   1250  DDGLLAHODAGRLARVLRPKVEGAWLHTLT   1280
```

```
Qy   175 --------------------------GGTMLITSGCAVAVNIIMGLITLHQSGHGH--------SHGTTNQQE 212
                                   ||||  |: :||   |   |: :|||                   ||  :||
Db  1131 LVAAPPEGNAAPVSLSAEGSYLVIGGLG-ALGLLVARWLVERGAGHLVLISRHGLPDREE 1189
Qy   213 ----ENPSVRAAFIHVIGDFMQSMGVLVAAYILYFKPEYKYVDPICTFVFSILVLGTTLT 268
             |||                     ||| |||||||  |  |||      |:||||  |
Db  1190 WGRDQPPEVRARIAAIEALEAQGARVTVAA------------------------------ 1219
Qy   269 ILRDVILVLMEGTPKGVDFTAVRDLLLSVEGVEALHSLHIWALTVAQPVLS--VHIA--- 323
         |||  |:  |:|   :|          ||:   ||| :  |: |    |:|   |:  |||
Db  1220 ------------------------VD------VADAEGMAAL------LAAVEPPLRGVHAAGLL 1249
Qy   324 ---IAQNTDAQAVLKTASSRLQGKFHFHVT 351
             ||  |||      ::|     ::| ||:|
Db  1250 DDGLIAHQDAGRLARVLRPKVEGAWVLHTLT 1280
```

RESULT 22
US-08-993-088A-7
; Sequence 7, Application US/08993088A
; Patent No. 6287855
; GENERAL INFORMATION:
;   APPLICANT: Tan, Carina
;   APPLICANT: Sullivan, Kathleen
;   TITLE OF INVENTION: GALANIN RECEPTOR GALR2 AND
;   TITLE OF INVENTION: NUCLEOTIDES ENCODING SAME
;   NUMBER OF SEQUENCES: 20
;   CORRESPONDENCE ADDRESS:
;   ADDRESSEE: Merck & Co., Inc.
;   STREET: P.O. Box 2000, 126 E. Lincoln Ave.
;   CITY: Rahway
;   STATE: NJ
;   COUNTRY: USA
;   ZIP: 07065-0900
;   COMPUTER READABLE FORM:
;   MEDIUM TYPE: Diskette
;   COMPUTER: IBM Compatible
;   OPERATING SYSTEM: Windows
;   SOFTWARE: FastSEQ for Windows Version 2.0b
;   CURRENT APPLICATION DATA:
;   APPLICATION NUMBER: US/08/993,088A
;   FILING DATE: 18-DEC-1997
;   CLASSIFICATION: 530
;   PRIOR APPLICATION DATA:
;   APPLICATION NUMBER: 60/033,851
;   FILING DATE: 27-DEC-1996
;   ATTORNEY/AGENT INFORMATION:
;   NAME: Heber, Sheldon O.
;   REGISTRATION NUMBER: 38,179
;   REFERENCE/DOCKET NUMBER: 19846
;   TELECOMMUNICATION INFORMATION:
;   TELEPHONE: 732-594-1958
;   TELEFAX: 732-594-4720
;   TELEX:
; INFORMATION FOR SEQ ID NO: 7:
;   SEQUENCE CHARACTERISTICS:
;   LENGTH: 387 amino acids
;   TYPE: amino acid
;   STRANDEDNESS: single
;   TOPOLOGY: linear
;   MOLECULE TYPE: protein
US-08-993-088A-7

Query Match   4.4%; Score 84; DB 4; Length 387;
Best Local Similarity   22.8%; Pred. No. 0.79;
Matches   46; Conservative 34; Mismatches 76; Indels 46; Gaps   8;

```
Qy   82 LLFMIGEVVGGYLAHSLAVMTDAAHLLTDFASMLISL-----FSLWMSSRPATKTMNFG 135
        |:||:||| |     |||||                ||:|          |||   |   |||
Db   34 LIFLVG-TVGNTLV--LAVLLRGGQAVSTTNLFILNLGVADLCFILCCVFQATIYTLDG 90
Qy  136 WQRAEILGALVS-----VLSIWVTGVLVYLAVERLISGDYEIDGGTMLITSGCAAV 188
        |:|:|:            |: ||  |   || |   |   :|  | ||   || |  :|
```

RESULT 23
US-08-993-424B-7
; Sequence 7, Application US/08993424B
; Patent No. 6337206
; GENERAL INFORMATION:
;   APPLICANT: Tan, Carina
;   APPLICANT: Kolakowski, Lee F., Jr.
;   TITLE OF INVENTION: MOUSE GALANIN RECEPTOR GALR2 AND
;   TITLE OF INVENTION: NUCLEOTIDES ENCODING SAME
;   NUMBER OF SEQUENCES: 18
;   CORRESPONDENCE ADDRESS:
;   ADDRESSEE: Merck & Co., Inc.
;   STREET: P.O. Box 2000, 126 E. Lincoln Ave.
;   CITY: Rahway
;   STATE: NJ
;   COUNTRY: USA
;   ZIP: 07065-0900
;   COMPUTER READABLE FORM:
;   MEDIUM TYPE: Diskette
;   COMPUTER: IBM Compatible
;   OPERATING SYSTEM: Windows
;   SOFTWARE: FastSEQ for Windows Version 2.0b
;   CURRENT APPLICATION DATA:
;   APPLICATION NUMBER: US/08/993,424B
;   FILING DATE: 18-DEC-1997
;   CLASSIFICATION: 536
;   PRIOR APPLICATION DATA:
;   APPLICATION NUMBER: 60/033,851
;   FILING DATE: 27-DEC-1996
;   ATTORNEY/AGENT INFORMATION:
;   NAME: Heber, Sheldon O.
;   REGISTRATION NUMBER: 38,179
;   REFERENCE/DOCKET NUMBER: 19846NP2
;   TELECOMMUNICATION INFORMATION:
;   TELEPHONE: 732-594-1958
;   TELEFAX: 732-594-4720
;   TELEX:
; INFORMATION FOR SEQ ID NO: 7:
;   SEQUENCE CHARACTERISTICS:
;   LENGTH: 387 amino acids
;   TYPE: amino acid
;   STRANDEDNESS: single
;   TOPOLOGY: linear
;   MOLECULE TYPE: protein
US-08-993-424B-7

Query Match   4.4%; Score 84; DB 4; Length 387;
Best Local Similarity   22.8%; Pred. No. 0.79;
Matches   46; Conservative 34; Mismatches 76; Indels 46; Gaps   8;

```
Qy   82 LLFMIGEVVGGYLAHSLAVMTDAAHLLTDFASMLISL-----FSLWMSSRPATKTMNFG 135
        |:||:||| |     |||||                ||:|          |||   |   |||
Db   34 LIFLVG-TVGNTLV--LAVLLRGGQAVSTTNLFILNLGVADLCFILCCVFQATIYTLDG 90
Qy  136 WQRAEILGALVS-----VLSIWVTGVLVYLAVERLISGDYEIDGGTMLITSGCAAV 188
        |:|:|:            |: ||  |   || |   |   :|  | ||   || |  :|
Db   91 W-----VFGSLLCKAVHFLIFLTMHASSFTLAAVSLDRYLAIRYPLHSRELRTPRNALAI 146
Qy  189 NIIMGLTLHQSGHGHSHGTTNQQEENPSVRAAFIHVIGDFMQSMGVLVAAYILYFKPEYK 248
        |: ||| |: |||    |||||||                          ||  ||||||
Db  147 GLIWGLSLLFSGPYLSYYRQSQ---------------------------LANLTVCHPAWSAPRRR 185
Qy  249 YVDPICFVFES-----ILVLGTT 266
         |||       |       |||||
Db  186 AMD-ICTFVFSILLPVLVLGLT 206
```

Qy   249 YVDPICTEVFS------ILVLGTT 266
          :||||||| ||||| |
Db   186 AMD-ICTFVFSYLLPVLVLGLT 206

RESULT 24
US-08-750-723A-2
Sequence 2, Application US/0875O723A
Patent No. 5981219
GENERAL INFORMATION:
;   APPLICANT: Flugge, Ulf-Ingo
;   APPLICANT: Weber, Andreas
;   APPLICANT: Fischer, Karsten
;   TITLE OF INVENTION: DNA MOLECULES WHICH CODE FOR A PLASTID
;   TITLE OF INVENTION: 2-OXOGLUTARATE/MALATE TRANSLOCATOR
;   NUMBER OF SEQUENCES: 2
;   CORRESPONDENCE ADDRESS:
;   ADDRESSEE: James F. Haley, Jr., Esq c/o FISH & NEAVE
;   STREET: 1251 Avenue of the Americas
;   CITY: New York
;   STATE: New York
;   COUNTRY: United States of America
;   ZIP: 10020
;   COMPUTER READABLE FORM:
;   MEDIUM TYPE: Floppy disk
;   COMPUTER: IBM PC compatible
;   OPERATING SYSTEM: PC-DOS/MS-DOS
;   SOFTWARE: PatentIn Release #1.0, Version #1.30 (EPO)
;   CURRENT APPLICATION DATA:
;   APPLICATION NUMBER: US/08/750,723A
;   FILING DATE:
;   CLASSIFICATION:
;   PRIOR APPLICATION DATA:
;   APPLICATION NUMBER: 435
;   FILING DATE:
;   APPLICATION NUMBER: DE P 44 20 782.4
;   FILING DATE: 15-JUN-1994
;   ATTORNEY/AGENT INFORMATION:
;   NAME: James F. Haley, Jr., Esq.
;   REGISTRATION NUMBER: 27,794
;   REFERENCE/DOCKET NUMBER: GFB-3
;   TELECOMMUNICATION INFORMATION:
;   TELEPHONE: (212) 596-9000
;   TELEFAX: (212) 596-9090
;   INFORMATION FOR SEQ ID NO: 2:
;   SEQUENCE CHARACTERISTICS:
;   LENGTH: 569 amino acids
;   TYPE: amino acid
;   TOPOLOGY: linear
;   MOLECULE TYPE: protein
US-08-750-723A-2

Query Match   4.4%; Score 84; DB 2; Length 569;
Best Local Similarity 21.2%; Pred. No. 1.4;
Matches   53; Conservative 34; Mismatches 67; Indels 96; Gaps 12;

Qy 107 LITDFASMLI--SLFSLWMSSRPATKTMNEG--------------WQRAEILGALVSVLSIWVV 154
        :!|!   :! ! | : | ||| ||:| |:| |:|
Db 273 MLTCFQTSVISSSMFLTAMAANPLSATLTFNTIGKAIGWMDWAKAAFVPGLVSL----IV 328

Qy 155 TGVLVY------------LAVERL----ISGDYEIDGGTMLITSG----------- 183
        :!:! |||:| !: |:|! :||||!!:|
Db 329 VPLLLYVVYPPEIKSSPDAPRLAKEKLDKMGPMTKNESIMAVTLLLTVGLWVFGGKLGVD 388

Qy 184 ------CAVAVNIIMGLT----------LHQSGHGHSHGTTNQQEENPS 216
        ::| :| !:| ||
Db 389 AVTAAILGLSVLLITGVVTWKECLAESVAWDTLTWFAALIAMAGYLNKYGLITWFSEN- 446

Qy 217 VRAAFIHVIG----DFMQSMGVLVAAYILYFKPEYKYVDPICTFVFSILVLGTTLTLRD 272
        :!|| :|| | | | | ! ||||
Db 447 ----VVKVVGGLGLSWQMSFGVLV---LLYFYSHY------FFASGAAHIGAMPTAFLS 492

Qy 273 VILVLMEGTP 282
        | !||
Db 493 VASAL--GTP 500

Query Match   4.4%; Score 84; DB 4; Length 569;
Best Local Similarity 21.2%; Pred. No. 1.4;
Matches   53; Conservative 34; Mismatches 67; Indels 96; Gaps 12;

RESULT 25
US-09-191-275-2
Sequence 2, Application US/09191275
Patent No. 6225526
GENERAL INFORMATION:
;   APPLICANT: Flugge, Ulf-Ingo
;   APPLICANT: Weber, Andreas
;   APPLICANT: Fischer, Karsten
;   TITLE OF INVENTION: DNA MOLECULES WHICH CODE FOR A PLASTID
;   TITLE OF INVENTION: 2-OXOGLUTARATE/MALATE TRANSLOCATOR
;   NUMBER OF SEQUENCES: 2
;   CORRESPONDENCE ADDRESS:
;   ADDRESSEE: James F. Haley, Jr., Esq c/o FISH & NEAVE
;   STREET: 1251 Avenue of the Americas
;   CITY: New York
;   STATE: New York
;   COUNTRY: United States of America
;   ZIP: 10020
;   COMPUTER READABLE FORM:
;   MEDIUM TYPE: Floppy disk
;   COMPUTER: IBM PC compatible
;   OPERATING SYSTEM: PC-DOS/MS-DOS
;   SOFTWARE: PatentIn Release #1.0, Version #1.30 (EPO)
;   CURRENT APPLICATION DATA:
;   APPLICATION NUMBER: US/09/191,275
;   FILING DATE:
;   CLASSIFICATION:
;   PRIOR APPLICATION DATA:
;   APPLICATION NUMBER: 08/750,723
;   FILING DATE:
;   ATTORNEY/AGENT INFORMATION:
;   NAME: James F. Haley, Jr., Esq.
;   REGISTRATION NUMBER: 27,794
;   REFERENCE/DOCKET NUMBER: GFB-3
;   TELECOMMUNICATION INFORMATION:
;   TELEPHONE: (212) 596-9000
;   TELEFAX: (212) 596-9090
;   INFORMATION FOR SEQ ID NO: 2:
;   SEQUENCE CHARACTERISTICS:
;   LENGTH: 569 amino acids
;   TYPE: amino acid
;   TOPOLOGY: linear
;   MOLECULE TYPE: protein
US-09-191-275-2

Query Match   4.4%; Score 84; DB 4; Length 569;
Best Local Similarity 21.2%; Pred. No. 1.4;
Matches   53; Conservative 34; Mismatches 67; Indels 96; Gaps 12;

Qy 107 LITDFASMLI--SLFSLWMSSRPATKTMNEG--------------WQRAEILGALVSVLSIWVV 154
        :!|!   :! ! | : | ||| ||:| |:| |:|
Db 273 MLTCFQTSVISSSMFLTAMAANPLSATLTFNTIGKAIGWMDWAKAAFVPGLVSL----IV 328

Qy 155 TGVLVY------------LAVERL----ISGDYEIDGGTMLITSG----------- 183
        :!:! |||:| !: |:|! :||||!!:|
Db 329 VPLLLYVVYPPEIKSSPDAPRLAKEKLDKMGPMTKNESIMAVTLLLTVGLWVFGGKLGVD 388

Qy 184 ------CAVAVNIIMGLT----------LHQSGHGHSHGTTNQQEENPS 216
        ::| :| !:| ||
Db 389 AVTAAILGLSVLLITGVVTWKECLAESVAWDTLTWFAALIAMAGYLNKYGLITWFSEN- 446

Qy 217 VRAAFIHVIG----DFMQSMGVLVAAYILYFKPEYKYVDPICTFVFSILVLGTTLTLRD 272
        :!|| :|| | | | | ! ||||
Db 447 ----VVKVVGGLGLSWQMSFGVLV.---LLYFYSHY------FFASGAAHIGAMPTAFLS 492

Qy 273 VILVLMEGTP 282
        | !||
Db 493 VASAL--GTP 500

RESULT 26

US-08-454-549-4
Sequence 4, Application US/08454549
Patent No. 5866324
GENERAL INFORMATION:
; APPLICANT: EPPLER, C. Mark
; APPLICANT: OZENBERGER, Bradley A.
; APPLICANT: HULMES, Jeffrey D.
; TITLE OF INVENTION: cDNA'S ENCODING PROTEINS CLOSELY RELATED
; TITLE OF INVENTION: TO OPIOID RECEPTORS
; NUMBER OF SEQUENCES: 13
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Darby & Darby, P.C.
; STREET: 805 Third Avenue
; CITY: New York
; STATE: New York
; COUNTRY: USA
; ZIP: 10022
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/454,549
; FILING DATE: 30-MAY-1995
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Robinson, Joseph R.
; REGISTRATION NUMBER: 33,448
; REFERENCE/DOCKET NUMBER: 0646/1A818-US5
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212) 527-7700
; TELEFAX: (212) 753-6237
; TELEX: 236687
INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 367 amino acids
; TYPE: amino acid
; STRANDEDNESS: not relevant
; TOPOLOGY: not relevant
; MOLECULE TYPE: protein
; ORIGINAL SOURCE:
; ORGANISM: Rat
US-08-454-549-4

Query Match   4.3%;   Score 83.5;   DB 2;   Length 367;
Best Local Similarity   20.6%;   Pred. No. 0.82;
Matches   81;   Conservative   59;   Mismatches 139;   Indels 115;   Gaps   20;

Qy   32 PLPRPGLDLQAIELAAQSNHHCHAQKGPDSHCDPKKG-KAQRQLYVA---------SAICLL 83
               ::|| :: |:.: |:.:   :||.|
Db   3 PVPSARAELQFSLLANVSDTFPSAFPSASANASGSPGARSASSLALAIATTALYSAVCAV 62

Qy   84 FMIGEVVGGYLAHSLAVMTDAAHLLTDFASMLISLFSLWMSSRPATKTMNF--------G 135
               :| |:   ::|:|| | :||:| |:
Db   63 GLLGNVL------------VMFGIVRYLKTATNIYIFNLALADALATSTLPFQSAKYLMET 113

Qy   136 WQRAEIL-GALVSV-------LSIWVVTGVLV--YLAVERLISG-DYEIDGGTMLI---- 180
               |:| |:| ::|::||:| ::||: ||
Db   114 WPFGELLCKAVLSIDYYNMFTSIFTLTMMSVDRYIAVCHPVKALDFRTPAKAKLINICIW 173

Qy   181 --TSGCAVANVIMGLTLHQSGHGHSHGTTNQQEENPSVRAAFIHVIGDFMQSMGVLVAA 238
                     :| |: : ||:|.|| ||||||||||||||||: :|| :
Db   174 VLASGVGVPI-MVMAVTQPRDG-------------------------------AVVC 198

Qy   239 YILYFKPEYKY--VDPICTFVFSILVLGTTLTILRDVILV----------LMEGT-PKGVDFTA 289
               ||:| ||:|| :||:|:||||||| | |:
Db   199 TLQFPSPSWNDTVTKICVFLFAFVVPILIITVCYGLMLLRLSVRLLSGSKEKDRSLRR 258

Qy   290 VRDLLLSVEG--VEALHSLHI---------NALTVAQ-----------PVLSVHIAIRAQNTDRAQAVLKTAS 338
               |:|::   :|:| : |:   :|:| | ::|:   :||
Db   259 ITRMVLVVVGAFVVCWAPIHIFVIWTLVDINRRDPLVVAALHICIA--------LGYAN 310

Qy   339 SRLQGKFHFHTVTIQIEDYSEDMKDC--QACQGP 370
               :| |: : ::|
Db   311 SSLN--------PVLYAFLDENFKRCFRQLCRAP 336

RESULT 27
US-08-454-552-4
Sequence 4, Application US/08454552
Patent No. 6005072
GENERAL INFORMATION:
; APPLICANT: EPPLER, C. Mark
; APPLICANT: OZENBERGER, Bradley A.
; APPLICANT: HULMES, Jeffrey D.
; TITLE OF INVENTION: cDNA'S ENCODING PROTEINS CLOSELY RELATED
; TITLE OF INVENTION: TO OPIOID RECEPTORS
; NUMBER OF SEQUENCES: 13
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Darby & Darby, P.C.
; STREET: 805 Third Avenue
; CITY: New York
; STATE: New York
; COUNTRY: USA
; ZIP: 10022
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/454,552
; FILING DATE: 30-MAY-1995
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Robinson, Joseph R.
; REGISTRATION NUMBER: 33,448
; REFERENCE/DOCKET NUMBER: 0646/1A818-US4
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212) 527-7700
; TELEFAX: (212) 753-6237
; TELEX: 236687
INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 367 amino acids
; TYPE: amino acid
; STRANDEDNESS: not relevant
; TOPOLOGY: not relevant
; MOLECULE TYPE: protein
; ORIGINAL SOURCE:
; ORGANISM: Rat
US-08-454-552-4

Query Match   4.3%;   Score 83.5;   DB 3;   Length 367;
Best Local Similarity   20.6%;   Pred. No. 0.82;
Matches   81;   Conservative   59;   Mismatches 139;   Indels 115;   Gaps   20;

Qy   32 PLPRPGLDLQAIELAAQSNHHCHAQKGPDSHCDPKKG-KAQRQLYVA---------SAICLL 83
               ::|| :: |:.: |:.:   :||.|
Db   3 PVPSARAELQFSLLANVSDTFPSAFPSASANASGSPGARSASSLALAIATTALYSAVCAV 62

Qy   84 FMIGEVVGGYLAHSLAVMTDAAHLLTDFASMLISLFSLWMSSRPATKTMNF--------G 135
               :| |:   ::|:|| | :||:| |:
Db   63 GLLGNVL------------VMFGIVRYLKTATNIYIFNLALADALATSTLPFQSAKYLMET 113

Qy   136 WQRAEIL-GALVSV-------LSIWVVTGVLV--YLAVERLISG-DYEIDGGTMLI---- 180
               |:| |:| ::|::||:| ::||: ||
Db   114 WPFGELLCKAVLSIDYYNMFTSIFTLTMMSVDRYIAVCHPVKALDFRTPAKAKLINICIW 173

Qy   181 --TSGCAVANVIMGLTLHQSGHGHSHGTTNQQEENPSVRAAFIHVIGDFMQSMGVLVAA 238
                     :| |: : ||:|.|| ||||||||||||||||: :|| :
Db   174 VLASGVGVPI-MVMAVTQPRDG-------------------------------AVVC 198

Qy   239 YILYFKPEYKY--VDPICTFVFSILVLGTTLTILRDVILV----------LMEGT-PKGVDFTA 289

RESULT 28
US-08-676-351-3
; Sequence 3, Application US/08676351C
; Patent No. 6046026
; GENERAL INFORMATION:
; APPLICANT: EPPLER, CECIL
; APPLICANT: OZENBERGER, BRADLEY
; APPLICANT: HULMES, JEFFREY
; TITLE OF INVENTION: cDNAS ENCODING PROTEINS CLOSELY RELATED
; TITLE OF INVENTION: TO OPIOID RECEPTORS
; FILE REFERENCE: 0646/1A818US1
; CURRENT APPLICATION NUMBER: US/08/676,351C
; CURRENT FILING DATE: 1996-09-12
; EARLIER APPLICATION NUMBER: PCT/US95/00939
; EARLIER FILING DATE: 1995-01-20
; EARLIER APPLICATION NUMBER: US 08/185,360
; EARLIER FILING DATE: 1994-01-21
; NUMBER OF SEQ ID NOS: 11
; SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 3
; TYPE: PRT
; ORGANISM: Rat
; LENGTH: 367

Query Match   4.3%; Score 83.5; DB 3; Length 367;
Best Local Similarity   20.6%; Pred. No. 0.82;
Matches   81; Conservative   59; Mismatches 139; Indels 115; Gaps 20;

```
Qy   32  PLPRPGLDLQAIELAAQSNHHCHAQKGPDSHCDPKKG--KAQRQLVA-------SAICLL  83
         |:|          :||     :  | |       :                 ||:
Db    3  PVPSARAELQFSLLANVSDTPPSAFPSASANASGSPGARSASSSLALAIAITALYSAVCAV  62

Qy   84  FMIGEVVGGYLAHSLAVMTDAAHLLTDFASMLISLFSLWMSSRPATKTMNF-------G  135
         ||:|:: | |: |             | |  | : |
Db   63  GLLGNVL-------VMFGIVRYTKLKTATNIYIFNLALADALATSTLPFQSAKYLMET  113

Qy  136  WQRAEIL-GALVSV-------LSIWVTGVLV--YLAVERLISG-DYEIDGGTMLI-----  180
         |      |       ||::    |              | |
Db  114  WPFGELLCKAVLSIDYYNMFTSIFTLTMMSVDRYIAVCHPVKALDFRTPAKARLINICIW  173

Qy  181  -TSGCAVAVNIIMGLTLHQSGHGHSHGTTNQQEENPSVRAAFIHVIGDFMQSMGVLVAA  238
         ||    |             |:  | |  : | |  |                   |
Db  174  VLASGVGVPI-MVMAVTQPRDG----------------------------------AVVC  198

Qy  239  YILYFKPEYKY--VDPICTFVFSILVLGTTLTLTLRDVILV--------LMEGT-PKGVDFTA  289
         |    |         |:|  |  |  |  |                  :: |   |
Db  199  TLQFPPSWYWDIVTKICVFLFAFVVPILIIITVCYGLMLLRLRSVRLLSGGSKEKDRSLRR  258

Qy  290  VRDLLLSVEG--VEALHSLHI----WALTVAQ-----PVLSVHIAIAQNTDAQAVLKTAS  338
         ||:|      |  |          | |:  |       :|
Db  259  ITRMVLVVVGAFVVCWAPIHIFVIVWLVDINRRDPLVVAALHLCIA------LGYAN  310

Qy  339  SRLQGKFHFTVTIQIEDYSEDMKDC--QACGGP  370
         | : |  |  | | |   :|
Db  311  SSLN------PVLYAFLDENFKRCFRQLCRAP  336
```

RESULT 29
US-08-188-275A-4
; Sequence 4, Application US/08188275A
; Patent No. 6258556

; GENERAL INFORMATION:
; APPLICANT: Uhl, George R.
; APPLICANT: Wang, Jia-Bel
; APPLICANT: Johnson, Peter S.
; APPLICANT: Persico, Antonio
; TITLE OF INVENTION: cDNA and Genomic Clones Encoding Human
; TITLE OF INVENTION: Mu Opiate Receptor and the Purified Gene Product
; NUMBER OF SEQUENCES: 12
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Birch, Stewart, Kolasch & Birch
; STREET: P.O. Box 747
; CITY: Falls Church
; STATE: Virginia
; COUNTRY: USA
; ZIP: 22040-3487
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/188,275A
; FILING DATE: 28-JAN-1994
; CLASSIFICATION: 530
; ATTORNEY/AGENT INFORMATION:
; NAME: Murphy Jr., Gerald M.
; REGISTRATION NUMBER: 28,977
; REFERENCE/DOCKET NUMBER: 1173-449P
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 703-241-1300
; TELEFAX: 703-241-2848
; TELEX: 248345
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 372 amino acids
; TYPE: amino acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; FRAGMENT TYPE: internal
; FEATURE:
; NAME/KEY: Protein
; LOCATION: 1..372
; OTHER INFORMATION: /label= protein
; OTHER INFORMATION: /note= "rat delta opiate receptor"
US-08-188-275A-4

Query Match   4.3%; Score 83.5; DB 4; Length 372;
Best Local Similarity   20.6%; Pred. No. 0.84;
Matches   81; Conservative   59; Mismatches 139; Indels 115; Gaps 20;

```
Qy   32  PLPRPGLDLQAIELAAQSNHHCHAQKGPDSHCDPKKG--KAQRQLVA-------SAICLL  83
         |:|          :||     :  | |       :                 ||:
Db    3  PVPSARAELQFSLLANVSDTPPSAFPSASANASGSPGARSASSSLALAIAITALYSAVCAV  62

Qy   84  FMIGEVVGGYLAHSLAVMTDAAHLLTDFASMLISLFSLWMSSRPATKTMNF-------G  135
         ||:|:: | |: |             | |  | : |
Db   63  GLLGNVL-------VMFGIVRYTKLKTATNIYIFNLALADALATSTLPFQSAKYLMET  113

Qy  136  WQRAEIL-GALVSV-------LSIWVTGVLV--YLAVERLISG-DYEIDGGTMLI----  180
         |      |       ||::    |              | |
Db  114  WPFGELLCKAVLSIDYYNMFTSIFTLTMMSVDRYIAVCHPVKALDFRTPAKARLINICIW  173

Qy  181  -TSGCAVAVNIIMGLTLHQSGHGHSHGTTNQQEENPSVRAAFIHVIGDFMQSMGVLVAA  238
         ||    |             |:  | |  : | |  |                   |
Db  174  VLASGVGVPI-MVMAVTQPRDG----------------------------------AVVC  198

Qy  239  YILYFKPEYKY--VDPICTFVFSILVLGTTLTLTLRDVILV--------LMEGT-PKGVDFTA  289
         |    |         |:|  |  |  |  |                  :: |   |
Db  199  TLQFPPSWYWDIVTKICVFLFAFVVPILIIITVCYGLMLLRLRSVRLLSGGSKEKDRSLRR  258

Qy  290  VRDLLLSVEG--VEALHSLHI----WALTVAQ-----PVLSVHIAIAQNTDAQAVLKTAS  338
         ||:|      |  |          | |:  |       :|
```

Left column

```
Db   259  ITRMVLVVVGAFVVCWAPIHIFVIVWTLVDINRRDPLVVAALHLCIA-------LGYAN 310

Qy   339  SRLQGKFHFHTVTIQIEDYSEDMKDC--QACQGP 370
                  | :|  |  |  |     |  |:|  |   |
Db   311  SSLN------PVLYAFLDENFKRCFRQLCRAP 336
```

RESULT 30
US-09-351-198-4
; Sequence 4, Application US/0351198
; Patent No. 6335168
; GENERAL INFORMATION:
; APPLICANT: Kreek, Mary J
; APPLICANT: Laforge, Karl S
; APPLICANT: Yu, Lei
; APPLICANT: Tischfield, Jay A.
; TITLE OF INVENTION: ALELES OFTHE HUMAN MU OPIOID RECEPTOR, DIAGNOSTIC
; TITLE OF INVENTION: METHODS OF USING SAID ALLELES, AND METHODS OF TREATMENT
; TITLE OF INVENTION: BASED THEREON
; FILE REFERENCE: 600-1-226N
; CURRENT APPLICATION NUMBER: US/09/351,198
; CURRENT FILING DATE: 1999-07-09
; EARLIER APPLICATION NUMBER: 60/092,402
; EARLIER FILING DATE: 1998-07-10
; NUMBER OF SEQ ID NOS: 7
; SOFTWARE: Patentin Ver. 2.0
;
; SEQ ID NO 4
; TYPE: PRT
; ORGANISM: Rattus norvegicus
US-09-351-198-4

Query Match   4.3%; Score 83.5; DB 4; Length 372;
Best Local Similarity 20.6%; Pred. No. 0.84;
Matches   81; Conservative 59; Mismatches 139; Indels 115; Gaps 20;

```
Qy    32  PLPRPGLDLQAIELAAQSNHHCHAQKGPDSHCDPKKG-KAQRQLYVA------SAICLL 83
              |:|| |             |       | :|   |:|   |           |:|:|  :
Db     3  PVPSARAELQFSLLANVSDTFPSAFPSASANASGSPGARSASSLAIAITALYSAVCAV 62

Qy    84  FMIGEVVGGYIAHSLAVMTDAAHLLTDFASMLISLFSLWMSSRPATKTMNF-------G 135
              |                      :          |:|: |             |
Db    63  GLIGNVL-----------VMFGIVRYTKLKTATNIYIFNLALADALATSTLPFQSAKYLMET 113

Qy   136  WQRAEIL-GALVSV-----LSIWVTGVLV--YLAVERLISG-DYEIDGGTMLI----- 180
              |:|       |  |:  :  |  |:|:|:|  |:     :|   |   :||
Db   114  WPFGELLCKAVLSIDYNMFTSIFTLTMMSVDRYIAVCHPVKALDFRTPAKAKLINICIW 173

Qy   181  --TSGCAVAVNIIMGLTLHQSGHGHSHGTTNQQEENPSVRAAFIHVIGDFMQSMGVLVAA 238
                  ||  | | | | |  | |  |   | :| |      |:  | :| |:
Db   174  VLASGVGVPI-MVMAVTQPRDG-----------------------------AVVC 198

Qy   239  YILYFKPEYKY--VDPICTFVFSILVLGTTLTILRDVILV-----LMEGT-PKGVDFTA 289
              ||:||:|   :| :|::| |:|   ||  |::|  |  :           ||  | |
Db   199  TLQFPSPSWYWDTVTKICVFLFAFVVPILIITVCGIMLLRLRSVRLLSGSKEKDRSLRR 258

Qy   290  VRDLLLSVEG--VEALHSLHI----WALTVAQ------PVLSVHIAIAQNTDAQAVLKTAS 338
              | : : :||         |:|         | |       | |:|  |  |
Db   259  ITRMVLVVVGAFVVCWAPIHIFVIVWTLVDINRRDPLVVAALHLCIA------LGYAN 310

Qy   339  SRLQGKFHFHTVTIQIEDYSEDMKDC--QACQGP 370
              | :|  |  |  |     |  |:|  |   |
Db   311  SSLN------PVLYAFLDENFKRCFRQLCRAP 336
```

Right column

Query Match   4.3%; Score 83.5; DB 4; Length 372;
Best Local Similarity 20.6%; Pred. No. 0.84;
Matches   81; Conservative 59; Mismatches 139; Indels 115; Gaps 20;

```
Qy    32  PLPRPGLDLQAIELAAQSNHHCHAQKGPDSHCDPKKG-KAQRQLYVA------SAICLL 83
              |:|| |             |       | :|   |:|   |           |:|:|  :
Db     3  PVPSARAELQFSLLANVSDTFPSAFPSASANASGSPGARSASSLAIAITALYSAVCAV 62

Qy    84  FMIGEVVGGYIAHSLAVMTDAAHLLTDFASMLISLFSLWMSSRPATKTMNF-------G 135
              |                      :          |:|: |             |
Db    63  GLIGNVL-----------VMFGIVRYTKLKTATNIYIFNLALADALATSTLPFQSAKYLMET 113

Qy   136  WQRAEIL-GALVSV-----LSIWVTGVLV--YLAVERLISG-DYEIDGGTMLI----- 180
              |:|       |  |:  :  |  |:|:|:|  |:     :|   |   :||
Db   114  WPFGELLCKAVLSIDYNMFTSIFTLTMMSVDRYIAVCHPVKALDFRTPAKAKLINICIW 173

Qy   181  --TSGCAVAVNIIMGLTLHQSGHGHSHGTTNQQEENPSVRAAFIHVIGDFMQSMGVLVAA 238
                  ||  | | | | |  | |  |   | :| |      |:  | :| |:
Db   174  VLASGVGVPI-MVMAVTQPRDG-----------------------------AVVC 198

Qy   239  YILYFKPEYKY--VDPICTFVFSILVLGTTLTILRDVILV-----LMEGT-PKGVDFTA 289
              ||:||:|   :| :|::| |:|   ||  |::|  |  :           ||  | |
Db   199  TLQFPSPSWYWDTVTKICVFLFAFVVPILIITVCGIMLLRLRSVRLLSGSKEKDRSLRR 258

Qy   290  VRDLLLSVEG--VEALHSLHI----WALTVAQ------PVLSVHIAIAQNTDAQAVLKTAS 338
              | : : :||         |:|         | |       | |:|  |  |
Db   259  ITRMVLVVVGAFVVCWAPIHIFVIVWTLVDINRRDPLVVAALHLCIA------LGYAN 310

Qy   339  SRLQGKFHFHTVTIQIEDYSEDMKDC--QACQGP 370
              | :|  |  |  |     |  |:|  |   |
Db   311  SSLN------PVLYAFLDENFKRCFRQLCRAP 336
```

RESULT 32
US-09-369-364A-22
; Sequence 22, Application US/0369364A
; Patent No. 6391610
; GENERAL INFORMATION:
; APPLICANT: Apte, Suneel
; APPLICANT: Hurskainen, Tiina L.
; APPLICANT: Hirohata, Satoshi
; TITLE OF INVENTION: Nucleic Acids Encoding Zinc Metalloproteases
; FILE REFERENCE: 26473/4007/10-30-00
; CURRENT APPLICATION NUMBER: US/09/369,364A
; CURRENT FILING DATE: 1999-08-06
; NUMBER OF SEQ ID NOS: 31
; SOFTWARE: Patentin Ver. 2.1
;
; SEQ ID NO 22
; TYPE: PRT
; LENGTH: 518
; ORGANISM: Homo sapiens ADAMTS-5
; FEATURE:
; NAME/KEY: MOD_RES
; LOCATION: (99)
; OTHER INFORMATION: Xaa = M
US-09-369-364A-22

Query Match   4.3%; Score 83.5; DB 4; Length 518;

RESULT 31
US-09-1113-426-4
; Sequence 4, Application US/09113426
; Patent No. 6337207
; GENERAL INFORMATION:
; APPLICANT: Kreek, Mary J
; APPLICANT: Laforge, Karl S
; APPLICANT: Yu, Lei

RESULT 33
US-09-122-126B-15
Sequence 15, Application US/09122126B
Patent No. 6451575
GENERAL INFORMATION:
; Patent No. 6451575
; GENERAL INFORMATION:
; APPLICANT: Bristol-Myers Squibb Company
; TITLE OF INVENTION: AGGRECAN DEGRADING METALLO PROTEASES
; FILE REFERENCE: DM66909
; CURRENT APPLICATION NUMBER: US/09/122,126B
; CURRENT FILING DATE: 1998-07-24
; NUMBER OF SEQ ID NOS: 21
; SOFTWARE: Patentin version 3.0
;
; SEQ ID NO 15
; LENGTH: 930
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-122-126B-15

Query Match   22.4%; Score 83.5; DB 4; Length 930;
Best Local Similarity   22.4%; Pred. No. 3.6;
Matches   50; Conservative   27; Mismatches   69; Indels   77; Gaps   10;

Qy   155 TGVLVLYLAVERLISGDYEIDGGTMLITSGCAVANVNIIMGLILHQSGHGHSGHTTNQQEEN 214
              |||: ||| ::|||:|| :|| | ||:|| ||||||||| ::|| |||||||||||
Db   357 TRFTAYLALKKK-NGEYLINGKYMISTSETIIDIN---GTVMNYSGWSHR-------- 402

Qy   215 PSVRAAFIHVIGDFMQSMGVLVAAYILYFKPEYKYVDPICTFVFSILVLGTTLTILRDV- 273
              |:|| ||| | ::| | |||
Db   403 -----------DDFLHGMGYSATKEILI-----------VQILATDPTKPLDVR 434

Qy   274 --ILVLMEGTPKGVDFTAVRDLLLSVEGVEALHSLHIWALTVAQP--VLSVHIAIAQNTD 329
             : ||| |:| |||
Db   435 YSFFVPKKSTPK----------VNSVTSHGSNKVGSHT-SQPQWVTGPWLACSRTCD 480

Qy   330 AQAVLKTASSRLQGKFHFHTVTIQIEDYSEDM-KDCQACQGPS 371
             ||||
Db   481 T--------------GWHTRTVQCQCDGNRKLAKGCPLSQRPS 508

; STREET: One Market Plaza, Steuart Tower, Suite 2000
; CITY: San Francisco
; STATE: California
; COUNTRY: USA
; ZIP: 94105
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/560,005
; FILING DATE:
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Dow, Karen B.
; REGISTRATION NUMBER: 29,684
; REFERENCE/DOCKET NUMBER: 2307K-0624000
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 415-326-2400
; TELEFAX: 415-326-2422
INFORMATION FOR SEQ ID NO: 6:
SEQUENCE CHARACTERISTICS:
; LENGTH: 942 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; FEATURE:
; NAME/KEY: Region
; LOCATION: 1..942
; OTHER INFORMATION: /note= "majptase"
US-08-560-005-6

Query Match   21.2%; Score 83; DB 3; Length 942;
Best Local Similarity   21.2%; Pred. No. 4.1;
Matches   51; Conservative   37; Mismatches   83; Indels   70; Gaps   12;

Qy   193 GLTLHQSGHGHSGHTTNQQEENPS------------VRAAFI------HVIGDFMQSMGGV---LV 236
             |:| || || |||||||||||| |: || |: || || :|
Db   220 GVPMDQSSRGQDKPESLQPRQNKSKSEITDMVRSSTIVSDKAHILS--MQKFGLRDTIV 277

Qy   237 AAYILYFKPEYKYVDPICTFVFSILVLGTTLTILRDVILVLMEGTPKGVDFTAV--RDLL 294
             ::|:|||::|: |||: |:|||||||||||||||: ::| ::| ||
Db   278 KSHLLQKEEDYTIIQNFRFFAGTYNVNGQS---PKECLRIWLSNGIQAPDVYCVGFQELL 334

Qy   295 LSVEGV------------EALHS------LHIWALTVAQPVLSVHIAIAQNTD 329
             ||: |||: ||||||||||||:::: :|
Db   335 LSKEAFFHDTPKEEEWFKRAVSEGLHPDAKYAKVKLIRLVGIMLLLYVKQEHAAYISEVE 394

Qy   330 AQAVLKTASSRLQGK------FHFHTVTI--------QIEDY----SEDMKD----CQAC 367
             |:| :||||||| |:|:||| ::||| :|:| ||:|
Db   395 AETVGTGIMGRMGNKGGYVAIRFQFHNTSICVVNSHLAAHIEEYERRNQDYKDICSRMQFC 454

Qy   368 Q 368
             |
Db   455 Q 455

RESULT 34
US-08-560-005-6
Sequence 6, Application US/08560005
Patent No. 6001354
GENERAL INFORMATION:
; APPLICANT: Pot, David A.
; APPLICANT: Williams, Lewis T.
; APPLICANT: Jefferson, Anne Bennett
; APPLICANT: Majerus, Philip W.
; TITLE OF INVENTION: No. 6001354el Grb2 Associating Protein and Nucleic
; TITLE OF INVENTION: Acids Encoding Therefor
; NUMBER OF SEQUENCES: 10
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Townsend and Townsend and Crew

RESULT 35
US-09-418-540-6
Sequence 6, Application US/09418540
Patent No. 6296848
GENERAL INFORMATION:
; APPLICANT: Pot, David A.
; APPLICANT: Williams, Lewis T.
; APPLICANT: Jefferson, Anne Bennett
; APPLICANT: Majerus, Philip W.
; TITLE OF INVENTION: No. 6296848el Grb2 Associating Protein and Nucleic
; TITLE OF INVENTION: Acids Encoding Therefor
; NUMBER OF SEQUENCES: 10
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Townsend and Townsend and Crew
; STREET: One Market Plaza, Steuart Tower, Suite 2000

CITY: San Francisco
STATE: California
COUNTRY: USA
ZIP: 94105
COMPUTER READABLE FORM:
  MEDIUM TYPE: Floppy disk
  COMPUTER: IBM PC compatible
  OPERATING SYSTEM: PC-DOS/MS-DOS
  SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
  APPLICATION NUMBER: US/09/418,540
  FILING DATE: 14-OCT-1999
  CLASSIFICATION:
PRIOR APPLICATION DATA:
  APPLICATION NUMBER: US 08/560,005
  FILING DATE: 17-NOV-1995
ATTORNEY/AGENT INFORMATION:
  NAME: Dow, Karen B.
  REGISTRATION NUMBER: 29,684
  REFERENCE/DOCKET NUMBER: 2307K-0624000
TELECOMMUNICATION INFORMATION:
  TELEPHONE: 415-326-2400
  TELEFAX: 415-326-2422
INFORMATION FOR SEQ ID NO: 6:
  SEQUENCE CHARACTERISTICS:
    LENGTH: 942 amino acids
    TYPE: amino acid
    TOPOLOGY: linear
  MOLECULE TYPE: protein
  FEATURE:
    NAME/KEY: Region
    LOCATION: 1..942
    OTHER INFORMATION: /note= "majptase"
US-09-418-540-6

Query Match 4.3%; Score 83; DB 4; Length 942;
Best Local Similarity 21.2%; Pred. No. 4.1;
Matches 51; Conservative 37; Mismatches 83; Indels 70; Gaps 12;

```
Qy  193 GLTLHQSGHGHSHSHGTTNQQEENPS-----VRAAFI------HVIGDFMQSMGV---LV 236
Db  220 GVPMDQSSRGQDKPESLQPRQNKSKSEITDMVRSSTITVSDKAHILS--MQKFGLRDTIV 277

Qy  237 AAYILYFKPEYKYVDPICTEVFSILVLGTTILTILRDVILVLMEGTPKGVDFTAV--RDLL 294
Db  278 KSHLLQKEEDTYIQNFRFFAGTYNVNGQS--PKECLRWLSNGIQAPDVYCVGFQELL 334

Qy  295 LSVEGV------------EALHS--------LHIWALTVAQPVLSVHIAIAQNTD 329
Db  335 LSKEAFFHDTPKEEEWFKAVSEGLHPDAKYAKVKLIRLVGIMLLIVKQEHAAYISEVE 394

Qy  330 AQAVLKTASSRLQGK--------FHFHTVTI----QIEDY---SEDMKD-----CQAC 367
Db  395 AETVGTGIGMGRMGNKGGVAIRFQFHNTSICVVNSHLAAHIEEYERNQDYKDICSRMOFC 454

Qy  368 Q 368
Db  455 Q 455
```

RESULT 36
US-08-362-512A-2
Sequence 2, Application US/08362512A
Patent No. 5719043
GENERAL INFORMATION:
  APPLICANT: FROMMER, Wolf-Bernd
  TITLE OF INVENTION: DNA SEQUENCES FOR AN AMINO ACID
  TITLE OF INVENTION: TRANSPORTER, PLASMIDS, BACTERIA, YEASTS AND PLANTS
  TITLE OF INVENTION: CONTAINING A TRANSPORTER AND THEIR USE
  NUMBER OF SEQUENCES: 4
  CORRESPONDENCE ADDRESS:
    ADDRESSEE: Ostrolenk, Faber, Gerb & Soffen
    STREET: 1180 Avenue of the Americas

CITY: New York
STATE: NY
COUNTRY: US
ZIP: 10036-8403
COMPUTER READABLE FORM:
  MEDIUM TYPE: Floppy disk
  COMPUTER: IBM PC compatible
  OPERATING SYSTEM: PC-DOS/MS-DOS
  SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
  APPLICATION NUMBER: US/08/362,512A
  FILING DATE: 05-JAN-1995
  CLASSIFICATION: 800
PRIOR APPLICATION DATA:
  APPLICATION NUMBER: US 08/007,636
  FILING DATE: 21-JAN-1993
ATTORNEY/AGENT INFORMATION:
  NAME: Mellman, Edward A.
  REGISTRATION NUMBER: 24,735
  REFERENCE/DOCKET NUMBER: P/951-107
TELECOMMUNICATION INFORMATION:
  TELEPHONE: (212) 382-0700
  TELEFAX: (212) 382-0888
  TELEX: 236925
INFORMATION FOR SEQ ID NO: 2:
  SEQUENCE CHARACTERISTICS:
    LENGTH: 485 amino acids
    TYPE: amino acid
    TOPOLOGY: linear
  MOLECULE TYPE: protein
US-08-362-512A-2

Query Match 4.3%; Score 82.5; DB 1; Length 485;
Best Local Similarity 18.1%; Pred. No. 1.7;
Matches 63; Conservative 60; Mismatches 97; Indels 129; Gaps 19;

```
Qy   85 MIGEVVGGYLAHSLAVMT-------DAAH----LLTDFASMLI-SLFSLWMSSRPATKT 131
Db  132 LIGVTVGYTITASISLVAVGKSNCFHDKGHTADCTISNYPYMAVFGIIQVILSOIP--- 187

Qy  132 MNFGWQRAEILGALVSVLSIWVVTGVLVLYLAVERLISGDYEIDGGTMLITSGGCAVAVNII 191
Db  188 -NF--HKILSFLISIMAAVMSFTYAT-IGIGLAIATVAGGKV---GKISM--TGTAVGVDVT 238

Qy  192 MGILTLHQS-------GHGHSHGTT----------NQQEENPSVRAA-------FIHVI- 225
Db  239 AAQKIWRSFQAVGDIAFAYAYATVLIELQDTLRSSPAENKAMKRASLVGVSTTTFYILC 298

Qy  226 ------GDFMQSMGV------------LVAAYILYFKPEYKYVDPIC- 254
Db  299 GCIGYAAFGNNAPGDFLTDFGFGFEPEFWLIDFANACIAVHLIGAYQVFAQPIFOFVERKCN 358

Qy  255 -------TFVFSI-----------LVLGTTLTILRDVILVLMEGTPKGVDFTAVR 291
Db  359 RNYPDNKFITSEYSVNVPFLGKFNISLFRLVWRTAYVVITTVVAMIFP------FFNAIL 412

Qy  292 DLLLSVEGVEALHSLHIWALTVAQPVLSVHIAIAQNTDAQAVLKTASSR 340
Db  413 GL-------IGAASFWPLTVVFPV-EMHI------AQTKIKKYSAR 444
```

RESULT 37
US-08-964-939-2
Sequence 2, Application US/08964939
Patent No. 6245970
GENERAL INFORMATION:
  APPLICANT: FROMMER, Wolf-Bernd
  TITLE OF INVENTION: DNA SEQUENCES FOR AN AMINO ACID
  TITLE OF INVENTION: TRANSPORTER, PLASMIDS, BACTERIA, YEASTS AND PLANTS
  TITLE OF INVENTION: CONTAINING A TRANSPORTER AND THEIR USE
  NUMBER OF SEQUENCES: 4
  CORRESPONDENCE ADDRESS:
    ADDRESSEE: Ostrolenk, Faber, Gerb & Soffen

```
;   STREET: 1180 Avenue of the Americas
;   CITY: New York
;   STATE: NY
;   COUNTRY: US
;   ZIP: 10036-8403
;   COMPUTER READABLE FORM:
;   MEDIUM TYPE: Floppy disk
;   COMPUTER: IBM PC compatible
;   OPERATING SYSTEM: PC-DOS/MS-DOS
;   SOFTWARE: PatentIn Release #1.0, Version #1.25
;   CURRENT APPLICATION DATA:
;   APPLICATION NUMBER: US/08/964,939
;   FILING DATE:
;   CLASSIFICATION: 800
;   PRIOR APPLICATION DATA:
;   APPLICATION NUMBER: US 08/362,512
;   FILING DATE: 05-JAN-1995
;   APPLICATION NUMBER: US 08/007,636
;   FILING DATE: 21-JAN-1993
;   ATTORNEY/AGENT INFORMATION:
;   NAME: Mellman, Edward A.
;   REGISTRATION NUMBER: 24,735
;   REFERENCE/DOCKET NUMBER: P/951-107
;   TELECOMMUNICATION INFORMATION:
;   TELEPHONE: (212) 382-0700
;   TELEFAX: (212) 382-0888
;   TELEX: 236925
;   INFORMATION FOR SEQ ID NO: 2:
;   SEQUENCE CHARACTERISTICS:
;   LENGTH: 485 amino acids
;   TYPE: amino acid
;   TOPOLOGY: linear
;   MOLECULE TYPE: protein
US-08-964-939-2
```

```
Query Match    4.3%;   Score 82.5;   DB 4;   Length 485;
Best Local Similarity   18.1%;   Pred. No. 1.7;
Matches   63;   Conservative   60;   Mismatches   97;   Indels   129;   Gaps   19;
```

```
Qy   85  MIGEVGGYLAHSLAVMT-------DAAH---LLTDFASMLI-SLFSLWMSSRPATKT 131
         :||| | |||:  |||:|:    |:   |:   || | |  :|||||||||||| |
Db  132  LIGVTVGYTITASISLVAVGKSNCFHDKGHTADCTISNYPYMAVFGIIQVILSQIP--- 187

Qy  132  MNFGWQRAEILGALVSVLSIWVVTGVLVYLAVERLISGDYEIDGGTMLITSGCAVANII 191
         |  |  | || |  | :|:  |  |  | |:  ||||||||||||||||||||||||
Db  188  -NF--HKLSFLSIMAAVMSFTYAT-IGIGLAIATVAGGKV--GKTSM--TGTAVGVDVT 238

Qy  192  MGLTLHQS------GHGHSHGTT-----NQQEENPSVRAA-------FIHVI- 225
          : :                 :||                        |  |
Db  239  AAQKIWRSFQAVGDIAFAYAYATVLIEIQDTLRSSPAENKAMKRASLVGVSTTFVVILC 298

Qy  226  ------GDFMQSMGV--------------LVAAYILYFKPEYKYVDPIC- 254
               |||       |               :|||||||||||||||||:
Db  299  GCIGYAAFGNNAPGDFLTDFGGFEPFWLIDFANACIAVHLLIGAYQVFAQPIFQFVEKKCN 358

Qy  255  ------TFVFSI-------LVLGTTLTILRDVILVLMEGTPKGVDFTAVR 291
              ::|||           |:  || ||:|||| : | |||  ||
Db  359  RNYPDNKFITSEYSVNVPFLGKFNISLFRLVWRTAYVVITVVVAMIFP-----FFNAIL 412

Qy  292  DILLSVEGVEALHSLHIWALTVAQPVLSVHIAIAQNTDAQAVLKTASSR 340
          : : :  |     | |  |  |    :           ||
Db  413  GL-------IGAASFWPLIVYFPV-EMHI------AQTKIKKYSAR 444
```

```
;   CURRENT APPLICATION NUMBER: US/09/134,001C
;   CURRENT FILING DATE: 1998-08-13
;   PRIOR APPLICATION NUMBER: US 60/064,964
;   PRIOR FILING DATE: 1997-11-08
;   PRIOR APPLICATION NUMBER: US 60/055,779
;   PRIOR FILING DATE: 1997-08-14
;   SEQ ID NO 3703
;   LENGTH: 496
;   NUMBER OF SEQ ID NOS: 5674
;   TYPE: PRT
;   ORGANISM: Staphylococcus epidermidis
US-09-134-001C-3703
```

RESULT 38

```
Query Match    4.3%;   Score 82.5;   DB 4;   Length 496;
Best Local Similarity   22.0%;   Pred. No. 1.7;
Matches   48;   Conservative   31;   Mismatches   56;   Indels   83;   Gaps   11;
```

```
Qy   75  YVASAICLLFMIGEVVGGYLAHSLAVMTDAAHLLTDFASMLISLFSLWMSSRPATKTMNF 134
         :|| : |   | : | |||||||  || | |||||| | |||||||||||||||||
Db  130  YIPATVLSIFLFTVVKMIESH------DIVSEMQLWVFLLVSLFG----------QF 171

Qy  135  GWQRAEILGALVSVLSIWVVTGVLVYLAVERLISGDYEIDGGTMLITSGCAVANIIM-- 192
         |  | || | :| | |  |  :| | | :||||||||||||||||||||||||||
Db  172  FLENLSIANSLI-----ILIGMVVYFFVKKRLS--YFLIVGFMLSCIG-----NIIMFL 218

Qy  193  -------GLTLHQSGHGHSHGTTNQQEENPSVRAAFIHVIGDFMQSMGVLVAAYILYF 243
                 |  |       | ||                 :       | | |
Db  219  NFNFELIKDGLNTHYS-ISDSHG---------------------------------VTLF 250

Qy  244  K--PEYKYVDPICTFVFSILVLGTTLTILRDVILVLME 279
         |    |||||||||  | :  :         :
Db  251  KLVPEYMFINQM----------------IILTVISIVSIVLLK 277
```

RESULT 39
US-09-134-001C-5457
```
;   Sequence 5457, Application US/09134001C
;   Patent No. 6380370
;   GENERAL INFORMATION:
;   APPLICANT: Lynn Doucette-Stamm et al
;   TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOC
;   TITLE OF INVENTION: EPIDERMIDIS FOR DIAGNOSTICS AND THERAPEUTICS
;   FILE REFERENCE: GTC-007
;   CURRENT APPLICATION NUMBER: US/09/134,001C
;   CURRENT FILING DATE: 1998-08-13
;   PRIOR APPLICATION NUMBER: US 60/064,964
;   PRIOR FILING DATE: 1997-11-08
;   PRIOR APPLICATION NUMBER: US 60/055,779
;   PRIOR FILING DATE: 1997-08-14
;   SEQ ID NO 5457
;   LENGTH: 524
;   NUMBER OF SEQ ID NOS: 5674
;   TYPE: PRT
;   ORGANISM: Staphylococcus epidermidis
US-09-134-001C-5457
```

```
Query Match    4.3%;   Score 82.5;   DB 4;   Length 524;
Best Local Similarity   19.2%;   Pred. No. 1.9;
Matches   59;   Conservative   62;   Mismatches  112;   Indels   75;   Gaps   15;
```

```
Qy   41  QAIELAAQSNHHCHAQGPDSHCDPKKGKAQROLYVASAICLL--FMIGEVVGGYLAHS 97
         ::|| |              |  |:       |  | |       || |   |||||
Db    3  ETIKMAKGDYQAHTEK----YHDKKSKSKSYKPWHIIISFILILILLLPTPAG------ 52

Qy   98  LAVMTDAAHLLTDFASMLISLFSLWMSSRPATKTMNFGWQRAEILGALVSVLSIWVVTGV 157
         ||  |||||||||||||| :||:|||||||||||   |  |  || | :| | |  :|
Db   53  LPVMAKAALAILAFAVV-----MVV----TEAVTYPVSATLIGLMILLGLSPVQDL 101

Qy  158  LVYLAVERLISGDYEIDGGTMLIT-------SGCAVANIIMGLTHQSGHGHSHGTTN 209
         | | | :|||||||||||||||         |||||||||| | |       | ||
Db  102  SEKLGNPK--SGDIILKGSDILGTNNALSHAFSGFSTSAVALVAAALFLA------VA 151

Qy  210  QQEENPSVRAAF--IHVIGDFMQSM---GVLVAAYILYFKPEYKYVDPICTFVFSILVLG 264
```

```
Db  152 MOETNLHKRLALLVLSIVGNKTRNIVIGAILVSIVLAFFVPS------------ATARAG 199
Qy  265 TTLTILRDVILVLMEGTPKGVDFTAVRD-LLLSVEGVEALHSLHIWALTV-----AQPVLS 319
Db  200 AVVPILLGMIAA------FNVSKDSRLASLLIITAVQAVSIWNIGIKTAAAQNIVA 249
Qy  320 VHIAIAQN 327
Db  250 INF-INQN 256
```

RESULT 40
US-08-484-840-2
Sequence 2, Application US/08484840
Patent No. 5716788
GENERAL INFORMATION:
   APPLICANT: MOSCOW, Jeffrey A.
   APPLICANT: COWAN, Kenneth H.
   APPLICANT: DIXON, Kathy
   APPLICANT: HE, Rui
   TITLE OF INVENTION: A GENE ENCODING A HUMAN REDUCED FOLATE
   TITLE OF INVENTION: CARRIER (RFC) AND METHODS FOR THE TREATMENT OF
   TITLE OF INVENTION: METHOTREXATE-RESISTANT, TRANSPORT-DEFICIENT CANCER CELLS
   NUMBER OF SEQUENCES: 6
   CORRESPONDENCE ADDRESS:
      ADDRESSEE: Foley & Lardner
      STREET: 3000 K Street, N.W., Suite 500
      CITY: Washington
      STATE: D.C.
      COUNTRY: USA
      ZIP: 20007-5109
   COMPUTER READABLE FORM:
      MEDIUM TYPE: Floppy disk
      COMPUTER: IBM PC compatible
      OPERATING SYSTEM: PC-DOS/MS-DOS
      SOFTWARE: PatentIn Release #1.0, Version #1.30
   CURRENT APPLICATION DATA:
      APPLICATION NUMBER: US/08/484,840
      FILING DATE: 07-JUN-1995
      CLASSIFICATION: 530
   ATTORNEY/AGENT INFORMATION:
      NAME: BENT, Stephen A.
      REGISTRATION NUMBER: 29,768
      REFERENCE/DOCKET NUMBER: 40399/324/NIHD
   TELECOMMUNICATION INFORMATION:
      TELEPHONE: (202)672-5300
      TELEFAX: (202)672-5399
      TELEX: 904136
   INFORMATION FOR SEQ ID NO: 2:
   SEQUENCE CHARACTERISTICS:
      LENGTH: 591 amino acids
      TYPE: amino acid
      TOPOLOGY: linear
   MOLECULE TYPE: protein
US-08-484-840-2

Query Match   4.3%; Score 82.5; DB 1; Length 591;
Best Local Similarity   21.6%; Pred. No. 2.3;
Matches 58; Conservative 33; Mismatches 74; Indels 103; Gaps 12;
```

```
Qy  91 GGYLAHSLAVM---TDAAHLLTDFASML----ISLFSLWMSSRPAIKTMNFGWQRAEILG 143
Db 235 GGKLGHALRVACGDSVLARMLRELGDSLRRPQLRLWSLW----------------------- 273
Qy 144 ALVSVLSIWVTGVLVYLAV-------------ERLISGDYEIDGGTML--ITSGCA 185
Db 274 -------WVFNSAGYLVYYVHILMNEVDPTTNSARVYNGAADA-ASTLLGAITSFAA 324
Qy 186 VAVNI------IMGLTLHQSGHGHSHGTTNQOEENPSVRAAFIHVIGDEMQSMGVLV 236
Db 325 GFVKIRWARWSKLLIAGVTATQAG---------------LVFLLAHTRHPSSIWLCY 366
```

```
Qy 237 AAYILYFKPEYKYVDPICTFVFS----------------ILVLGTT---LTILRDVILVLMEGTPKGV 285
Db 367 AAFVL-FRGSYQFLVPIATFQIASSLSKELCALVFGVNFFAATIVKTIITFIV------ 418
Qy 286 DFTAVRDILLLSVEGVEALHSLHIWALTV 313
Db 419 --SDVRGLGLPVRKQFQLYSVYFLILSI 444
```

Search completed: March 18, 2003, 15:42:48
Job time : 37 secs

OM protein - protein search, using sw model

GenCore version 5.1.4_p5_4578
Copyright (c) 1993 - 2003 Compugen Ltd.

Run on: March 18, 2003, 15:40:11 ; Search time 33 Seconds
(without alignments)
2322.713 Million cell updates/sec

Title: 1923
Perfect score: US-09-691-219-2
Sequence: 1 MEAKEKQHLLDARPAIRSYT.........QIEDYSEDMKDCQACQGPSD 372

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 671580 seqs, 206047115 residues

Total number of hits satisfying chosen parameters: 671580

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 120 summaries

Database : SPTREMBL_21:*
1: sp_archea:*
2: sp_bacteria:*
3: sp_fungi:*
4: sp_human:*
5: sp_invertebrate:*
6: sp_mammal:*
7: sp_mhc:*
8: sp_organelle:*
9: sp_phage:*
10: sp_plant:*
11: sp_rodent:*
12: sp_virus:*
13: sp_vertebrate:*
14: sp_unclassified:*
15: sp_rvirus:*
16: sp_bacteriap:*
17: sp_archeap:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result NO.	Score	% Query Match	Length	DB	ID	Description
1	1634.5	85.0	323	4	Q9BRI3	Q9bri3 homo sapien
2	972.5	50.6	388	4	Q8TC03	Q8tc03 homo sapien
3	892.5	46.4	320	4	Q8TCL3	Q8tcl3 homo sapien
4	878.5	45.7	472	5	Q9V471	Q9v471 drosophila
5	838.5	43.6	382	5	Q22541	Q22541 caenorhabdi
6	811.5	42.2	1677	5	Q9VKA3	Q9vka3 drosophila
7	722.5	37.6	429	4	Q8TC39	Q8tc39 homo sapien
8	709.5	36.9	410	5	Q45923	Q45923 caenorhabdi
9	702	36.5	391	5	Q45922	Q45922 caenorhabdi
10	649.5	33.8	398	10	O81036	O81036 arabidopsis
11	649	33.7	360	10	Q94AZ9	Q94az9 thlaspi goe
12	643.5	33.5	398	10	Q9ZT63	Q9zt63 arabidopsis
13	642	33.4	378	10	Q9LXS1	Q9lxs1 arabidopsis
14	639	33.2	396	10	Q93XE9	Q93xe9 thlaspi cae
15	638	33.2	392	10	Q94B01	Q94b01 thlaspi goe
16	633	32.9	392	10	Q94B00	Q94b00 thlaspi goe
17	631	32.8	421	10	Q947R8	Q947r8 eucalyptus
18	605.5	31.5	498	5	Q8T0G1	Q8t0g1 drosophila
19	599	31.1	334	10	Q9M271	Q9m271 arabidopsis
20	565	29.4	452	5	Q9V9I1	Q9v9i1 drosophila
21	500.5	26.0	385	10	Q9ZW23	Q9zw23 arabidopsis
22	432	22.5	334	5	Q8SQT3	Q8sqt3 encephalito.
23	421.5	21.9	321	16	Q9PF12	Q9pf12 xylella fas
24	421.5	21.9	312	16	Q9X7Q7	Q9x7q7 streptococc
25	414.5	21.7	303	16	Q8Y480	Q8y480 listeria mo
26	410.5	21.3	303	16	Q927R8	Q927r8 listeria in
27	399.5	21.3	316	2	Q9R494	Q9r494 bacillus st
28	399.5	20.8	319	16	Q99SB4	Q99sb4 staphylococ
29	397.5	20.7	311	16	Q99X53	Q99x53 staphylococ
30	393.5	20.5	301	16	O07084	O07084 bacillus su
31	393.5	20.3	317	2	Q8YFT2	Q8yft2 brucella me
32	391	20.3	325	16	O05214	O05214 bacillus st
33	390	20.3	325	2	Q92NF5	Q92nf5 staphylococ
34	388.5	20.2	295	2	P71023	P71023 bacillus su
35	374.5	19.5	361	16	Q9ABC7	Q9abc7 caulobacter
36	372	19.3	299	16	Q916A3	Q916a3 pseudomonas
37	369	19.2	326	2	Q9I6A3	Q9i6a3 pseudomonas
38	366	19.0	449	5	O85143	O85143 staphylococ
39	366	18.7	365	5	Q9VZR4	Q9vzr4 drosophila
40	357	18.6	429	3	Q98H85	Q98h85 rhizobium 1
41	351.5	18.3	507	4	O06808	O06808 saccharomyc
42	351	18.3	325	16	Q9BZF6	Q9bzf6 homo sapien
43	346.5	18.0	308	16	Q9RVC7	Q9rvc7 deinococcus
44	336.5	17.5	303	16	O67168	O67168 aquifex aeo
45	334.5	17.4	284	16	Q97G32	Q97g32 clostridium
46	334	17.4	300	10	Q9WZ08	Q9wz08 thermotoga
47	332	17.3	304	10	Q9S103	Q9s103 arabidopsis
48	319.5	16.6	295	17	Q8ZSA1	Q8zsa1 anabaena sp
49	314	16.3	387	2	O8U3B2	O8u3b2 pyrococcus
50	313	16.3	291	17	O13918	O13918 schizosacch
51	309	16.1	316	16	Q8TRQ7	Q8trq7 methanosarc
52	302	15.7	146	4	Q9PND2	Q9pnd2 campylobact
53	296.5	15.4	295	17	Q961T7	Q961t7 homo sapien
54	290	15.1	775	5	O58637	O58637 pyrococcus
55	282.5	14.7	492	11	O8WQB6	O8wqb6 caenorhabdi
56	282.5	14.7	761	11	Q9JKN2	Q9jkn2 mus musculu
57	281.5	14.6	594	4	Q8R4H9	Q8r4h9 mus musculu
58	281.5	14.6	765	4	Q9H9X0	Q9h9x0 homo sapien
59	280.5	14.6	360	4	Q8TAD4	Q8tad4 homo sapien
60	277	14.4	692	4	Q9BY48	Q9by48 homo sapien
61	274	14.2	294	2	Q9H9H1	Q9h9h1 homo sapien
62	266	13.8	561	5	Q9RMF3	Q9rmf3 zymomonas m
63	263	13.7	519	5	Q95QW4	Q95qw4 caenorhabdi
64	262	13.6	366	5	O18009	O18009 caenorhabdi
65	248	12.9	523	4	Q9VGS1	Q9vgs1 drosophila
66	246	12.8	378	11	Q8TDG3	Q8tdg3 homo sapien
67	237	12.3	376	4	Q9JKN1	Q9jkn1 mus musculu
68	233	12.1	304	16	Q8TCH2	Q8tch2 homo sapien
69	230.5	12.0	338	16	Q9CF74	Q9cf74 lactococcus
70	229	11.9	299	5	O8UG24	O8ug24 agrobacteri
71	226.5	11.8	281	16	Q97NZ2	Q97nz2 streptococc
72	226	11.7	287	2	Q8YT81	Q8yt81 anabaena sp
73	218.5	11.4	271	17	Q9F8C5	Q9f8c5 streptococc
74	206	10.7	326	16	Q97WH2	Q97wh2 pseudomonas
75	204.5	10.6	295	16	Q91447	Q91447 homo sapien
76	201.5	10.5	288	16	Q9BTR6	Q9btr6 thermoanaer
77	200.5	10.4	513	5	Q8RDD5	Q8rdd5 drosophila
78	195.5	10.2	513	5	O8SYV5	O8syv5 drosophila
79	195.5	10.1	195	5	Q9VPE3	Q9vpe3 drosophila
80	191.5	10.0	740	3	Q9CLA2	Q9cla2 pasteurella
81	190	9.9	195	16	O03455	O03455 saccharomyc
82	187.5	9.9	291	5	Q9A0B8	Q9a0b8 streptococc
83	187	9.8	291	16	O8ZUM0	O8zum0 pyrobaculum
84	186.5	9.7	276	17	O69597	O69597 mycobacteri
85	186	9.7	240	16	O9NPW0	O9npw0 homo sapien
86	184	9.6	256	2	Q9CCX3	Q9ccx3 mycobacteri
87	184	9.6	268	2	O97K27	O97k27 clostridium
88	181.5	9.4	292	16	Q92P03	Q92p03 rhizobium m
89	180	9.1	321	3	O9HGG3	O9hgg3 schizosacch
90	172	8.9	164	10	P93574	P93574 solanum tub

```
90  168.5  8.8  290   16  P96610      Db   91  -----------------------------------------EILGALVSVLSIWVVTGVLVYLAVERLISGDYEIDGGTMLI  131
91         8.8  461    4  Q96NC3
92         8.5  393   16  Q8XMT3      Qy  181  TSGCAVAVNIIMGLTLHQSGHGHSHGTTNQQEENPSVRAAFIHVIGDPMQSMGVLVAAYI  240
93         8.5  298   17  027915           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
94         8.4  449   17  O08970      Db  132  TSGCAVAVNIIMGLTLHQSGHGHSHGTTNQQEENPSVRAAFIHVIGDPMQSMGVLVAAYI  191
95  158.5  8.2  300   11  Q8YT27
96  158.5  8.2  460   11  Q8R4Z2      Qy  241  LYFKPEYKYVDPICTFVFSILVLGTTLILTILRDVILVLMEGTPKGVDFTAVRDLLLSVEGV  300
97         8.2  317   16  Q9UZ4            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
98  157.5  8.2  483   10  082250      Db  192  LYFKPEYKYVDPICTFVFSILVLGTTLILTILRDVILVLMEGTPKGVDFTAVRDLLLSVEGV  251
99         8.1  288   16  Q99RL3
100 156    8.1  291   16  Q8Y4L4      Qy  301  EALHSLHIWALTVAQPVLSVHIAIAQNTDAQAVLKTASSRLQGKFHFHFTVTIQIEDYSED  360
101 156    8.1  384   17  030282           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
102 156    8.1  428   17  Q23281      Db  252  EALHSLHIWALTVAQPVLSVHIAIAQNTDAQAVLKTASSRLQGKFHFHFTVTIQIEDYSED  311
103 155    8.1  341   17  Q8TKN5
104 154    8.0  266   17  Q96YR3      Qy  361  MKDCQACQGPSD  372
105 154    8.0  302   16  Q9KNP1           ||||||||||||
106 153.5  8.0  306   16  Q8XVQ7      Db  312  MKDCQACQGPSD  323
107 151    7.9  331   17  Q8TT22
108 149    7.7  289   16  Q8Y545
109 147.5  7.7  299   17  Q9HP06
110 147.5  7.7  344    2  Q9RH59
111 145    7.7  291   16  Q92815
112 145    7.5  372   16  Q8REY6
113 144.5  7.5  291   16  Q8REY6
114 143.5  7.4  298   16  P78888
115 142.5  7.4  356    3  P78888
116 142    7.4  403   16  Q97LF5
117 141    7.3  290    2  P72068
118 139.5  7.3  311   16  Q8tt88
119 139    7.2  289   16  Q929E4
120 139    7.2  407   17  027240
121 137    7.1  338    2  085699
```

```
RESULT 1

ID    Q9BRI3
AC    Q9BRI3;
DT    01-JUN-2001 (TrEMBLrel. 17, Created)
DT    01-JUN-2001 (TrEMBLrel. 21, Last sequence update)
DT    01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE    Similar to zinc transporter 2.
OS    Homo sapiens (Human).
OC    Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC    Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX    NCBI_TaxID=9606;
RN    [1]
RP    SEQUENCE FROM N.A.
RC    TISSUE=OVARY;
RA    Strausberg R.;
RL    Submitted (APR-2001) to the EMBL/GenBank/DDBJ databases.
DR    EMBL; BC006251; AAH06251.1; -.
DR    InterPro; IPR002524; Cation_efflux.
DR    Pfam; PF01545; Cation_efflux; 1.
DR    TIGRFAMs; TIGR01297; CDF; 1.
SQ    SEQUENCE  323 AA;  35177 MW;  F487412DEC7B38CB CRC64;
```

```
PRELIMINARY;          PRT;          323 AA.
```

Query Match   85.0%; Score 1634.5; DB 4; Length 323;
Best Local Similarity 86.8%; Pred. No. 1.6e-132;
Matches 323; Conservative 0; Mismatches 0; Indels 49; Gaps 1;

ALIGNMENTS

```
Qy    1  MEAKEKQHLLDARPAIRSYTGSLWQEGAGWIPLPRPGCLDLQAIELAAQSNHHCHAQKGPD   60
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1  MEAKEKQHLLDARPAIRSYTGSLWQEGAGWIPLPRPGCLDLQAIELAAQSNHHCHAQKGPD   60

Qy   61  SHCDPKKGKAQRQLYASAICLLFMIGEVVGGYLAHSLAVMTDAAHLLTDFASMLISLFS   120
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   61  SHCDPKKGKAQRQLYASAICLLFMIGEVV                                 90

Qy  121  LWMSSRPATKTMNFGWQRAEILGALVSVLSIWVVTGVLVYLAVERLISGDYEIDGGTMLI  180
         ||||||||||
```

```
RESULT 2

ID    Q8TCO3
AC    Q8TCO3;
DT    01-JUN-2002 (TrEMBLrel. 21, Created)
DT    01-JUN-2002 (TrEMBLrel. 21, Last sequence update)
DT    01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE    Solute carrier family 30 (zinc transporter), member 3.
OS    Homo sapiens (Human).
OC    Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC    Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX    NCBI_TaxID=9606;
RN    [1]
RP    SEQUENCE FROM N.A.
RC    TISSUE=TESTIS;
RA    Strausberg R.;
RL    Submitted (APR-2002) to the EMBL/GenBank/DDBJ databases.
DR    EMBL; BC028358; AAH28358.1; -.
SQ    SEQUENCE  388 AA;  41945 MW;  D73DE65F46860FD8 CRC64;
```

```
PRELIMINARY;          PRT;          388 AA.
```

Query Match   50.6%; Score 972.5; DB 4; Length 388;
Best Local Similarity 53.7%; Pred. No. 1.8e-75;
Matches 189; Conservative 50; Mismatches 94; Indels 19; Gaps 4;

```
Qy   32  PLPRPGCLDLQAIELAAQSNHHCHAQKGPDSHCDPKKGKAQRQLYASAICLLFMIGEVVG   91
         |:|    |    |    :  :|         |::|:||:||||||||||||
Db   40  PLPE----ESKPVEMPF---HHCHRDPLPPPGLTPERLHARRQLYAACAVCFVFMAGEVVG   93

Qy   92  GYLAHSLAVMTDAAHLLTDFASMLISLFSLWMSSRPATKTMNFGWQRAEILGALVSVLSI  151
         |||||| ||:|||:|:|||:||:|| |||:|   |      ||      ||:||||:|:|:
Db   40  GYLAHSLAIMTDAAHLLADVGSMMGSLFSLWLSRPATRTMTFGWHRSETLGALASVVSL  153

Qy  152  WVVTGVLVYLAVERLISGDYEIDGGTMLIITSGCAVAVNIIMGLTLHQSGHGHSHGTTNQ  211
         |:|||:||:|||||| ||:||:||||| |||:|:|   |||||  |||||| ||   |
Db  154  WMVTGILLYLAFVRLIHSDYHIEGGAMLITASIAVCANILMAFVLHQAGPPHSGSRGAE  213

Qy  212  --------EENP----------SVRAAFIHVIGDPMQSMGVLVAAYILYFKPEYKYVDPICTFVF  258
                 ::||             | || |   |::|: || | | : || |::||
Db  214  YAPLEEGPEEPLPLGNTSVRAAFVHVLGDLLQSFGVLAASILIYFKPQYKAADPISTFLF  273

Qy  259  SILVLGTTLTLILRDVILVLMEGTPKGVDFTAVRDLLLSVEGVEALHSLHIWALTVAQPVL  318
         :| ||: |  |: |||| || |:  |   : | || || | |||||||||||||||||||
Db  274  SICALGSTAPTLRDVLRILMEGTPRNVGPEPVRDTLLSVPGVRATHELHLWALTLTYHVA  333

Qy  319  SVHIAIAQNTDAQAVLKTASSRLQGKFHFHFTVTVIQIEDYSEDMKDCQACQGP  370
         | | || |: |   ||||||:|||||:||:||||:|||||| ||||||||||
Db  334  SAHLAIDSTADPEAVLAEASSRLYSRFGFSSCTLQVEQYQPEMAQCCLRCQEP  385
```

```
RESULT 3

ID    Q8TCL3
AC    Q8TCL3;
DT    01-JUN-2002 (TrEMBLrel. 21, Created)
```

```
PRELIMINARY;          PRT;          320 AA.
```

```
DT   01-JUN-2002 (TrEMBLrel. 21, Last sequence update)
DT   01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE   Hypothetical 35.1 kDa protein.
GN   DKFZp564F1062.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=BRAIN;
RA   Blum H., Bauersachs S., Mewes H.W., Gassenhuber J., Wiemann S.;
RL   Submitted (MAR-2002) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AL713790; CAD28545.1; -.
KW   Hypothetical protein.
SQ   SEQUENCE   320 AA;   35053 MW;   5D7325EBD1758EC9 CRC64;

Query Match             46.4%;  Score 892.5;  DB 4;  Length 320;
Best Local Similarity   52.8%;  Pred. No. 1.1e-68;
Matches  170;  Conservative   60;  Mismatches   89;  Indels    3;  Gaps    1;

Qy   51 HHCHAQKGPDSHCDPKKGKAQROLYASAICLLFMIGEVVGGYLAHSLAVMIDAAHLLTD 110
         :!!!:      :!.!:!:!::!:!   !:::!:!!:!: !!:!:!!!:! !!!!:!!!!!!
Db    2 YHCHSGSKPTEKGANEYAYAKWKLCSASAICFFMIAEVVGGHIAGSLAVVDAAHLLID 61

Qy  111 FASMLISFLSLWMSSRPATKTMNFGWQORAEILGALVSVLSIWWTGVIYLVLAVERLISGD 170
         !:!!!!!!!:!:!!! !!!!!:  :! !!:!:! !!!!!!!!!!!!!  !!!!:! :!:
Db   62 LTSFLLSLFSLWLSSKPPSKRLTFGWHRAELIGALLSILCIWVVTGVIVLYLACERLLYPD 121

Qy  171 YEIDGGTMLITSGCAVANVIIMGLTLHQSGHGHSHGTTNQQEENPSVRAAFIHVIGDEMQ 230
         !:!:! :!!!!:!  !:!:!   !!!!  !:! :! : !!!!!!!!! :!:! !! !
Db  122 YQIQATVMIIVSSCAVAANIVLTVVLHQRCLGHNH---KEVQANASVRAAFVHIAPGDLFQ 178

Qy  231 SMGVLVAAYILYFKPEYKVVDPICTFVFSILVLGTTLTILRDVILVLVLMEGTPKGVDFTAV 290
         !! !!!:!:!!:!!!:!!!!! :!!:!:!!!! :!::! !:!:! : !!!!:!!!!:! !:
Db  179 SISVLISALIIYFKPEYKIADPICTFIPSILVLASTITILKDFSILLMEGVRKSLNYSGV 238

Qy  291 RDLLLSVEGVEALHSLHIWALTVAQPVLSVHIAIAQNTDAQAVLKTASSRLQGKFHFHTV 350
         :!:!!!!!:!! !!!!!!!!:!!!  !!!::!! :!!! !: !!:: !!:!: !:!!! !::
Db  239 KELILAVDGVLSVHCLHIWSLTMNQVLSAHVATAASRDSQVVRREIAKALSKSFTMHSL 298

Qy  351 TIQIEDYSEDMKDCQACQGPSD 372
         !:!:! !:!!:!! :! ! !!!
Db  299 TIOMESPVDQDPDCLFCEDPCD 320
```

```
RESULT 4
ID   Q9V471     PRELIMINARY;     PRT;   472 AA.
AC   Q9V471;
DT   01-MAY-2000 (TrEMBLrel. 13, Created)
DT   01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT   01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE   BG:DS07295.1 protein (RE54080p).
GN   BG:DS07295.1 OR CG3994.
OS   Drosophila melanogaster (Fruit fly).
OC   Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC   Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC   Ephydroidea; Drosophilidae; Drosophila.
OX   NCBI_TaxID=7227;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=BERKELEY;
RX   MEDLINE=20196006; PubMed=10731132;
RA   Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA   Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
RA   George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA   Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
RA   Brandon R.C., Rogers Y.-H.C., Blazej R.G., Champe M., Pfeiffer B.D.,
RA   Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
RA   Abril J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
RA   Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA   Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
RA   Borkova D., Botchan M.R., Bouck J., Brokstein P., Brottier P.,
RA   Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA   Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
RA   de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA   Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA   Durbin K.J., Evangelista C.C., Ferraz C., Ferriera S., Fleischmann W.,
RA   Fosler C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,
RA   Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA   Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
RA   Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
RA   Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA   Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA   Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA   Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
RA   Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
RA   Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA   Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacleb J.M.,
RA   Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA   Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
RA   Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
RA   Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
RA   Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA   Wang Z.-Y., Wassarman D.A., Weinstock G.M., Weissenbach J.,
RA   Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
RA   Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA   Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
RA   Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
RT   "The genome sequence of Drosophila melanogaster.";
RL   Science 287:2185-2195(2000).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=BERKELEY;
RX   MEDLINE=99403001; PubMed=10471707;
RA   Ashburner M., Misra S., Roote J., Lewis S.E., Blazej R., Davis T.,
RA   Doyle C., Galle R., George R., Harris N., Hartzell G., Harvey D.,
RA   Hong L., Houston K., Hoskins R., Johnson G., Martin C., Moshrefi A.,
RA   Palazzolo M., Reese M.G., Spradling A., Tsang G., Wan K., Whitelaw K.,
RA   Celniker S., Rubin G.M.;
RT   "An exploration of the sequence of a 2.9-Mb region of the genome of
RT   Drosophila melanogaster: the Adh region.";
RL   Genetics 153:179-219(1999).
RN   [3]
RP   SEQUENCE FROM N.A.
RC   STRAIN=BERKELEY;
RA   Celniker S.E., Agbayani A., Arcaina T.T., Baxter E., Blazej R.G.,
RA   Butenhoff C., Champe M., Chavez C., Chew M., Ciesiolka L., Doyle C.M.,
RA   Farfan D.E., Galle R., George R.A., Harris N.L., Hoskins R.A.,
RA   Houston K.A., Hummasti S.R., Karra K., Kearney L., Kim E., Lee B.,
RA   Lewis S., Li P., Lomotan M.A., Mazda P., Moshrefi A.R., Moshrefi M.,
RA   Nixon K., Pacleb J.M., Park S., Pfeiffer B., Poon L., Sequeira A.,
RA   Sethi H., Snir E., Svirskas R.R., Wan K.H., Weinburg T., Zhang R.,
RA   Zieran L.L., Rubin G.M.;
RL   Submitted (MAR-2000) to the EMBL/GenBank/DDBJ databases.
RN   [4]
RP   SEQUENCE FROM N.A.
RC   STRAIN=BERKELEY;
RA   Stapleton M., Brokstein P., Hong L., Agbayani A., Carlson J.,
RA   Champe M., Chavez C., Dorsett V., Dresnek D., Farfan D., Frise E.,
RA   George R., Gonzalez M., Guarin H., Kronmiller B., Li P., Liao G.,
RA   Miranda A., Mungall C.J., Nunoo J., Pacleb J., Paragas V., Park S.,
RA   Patel S., Phouanenavong S., Wan K., Yu C., Lewis S.E., Rubin G.M.,
RA   Celniker S.;
RL   Submitted (DEC-2001) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AE003646; AAF53443.1; -.
DR   EMBL; AE003412; AAF44926.1; -.
DR   EMBL; AY071460; AAL49082.1; -.
DR   FlyBase; FBgn0028516; BG:DS07295.1.
DR   InterPro; IPR002524; Cation_efflux.
DR   Pfam; PF01545; Cation_efflux; 1.
DR   TIGRFAMs; TIGR01297; CDF; 1.
SQ   SEQUENCE   472 AA;   51036 MW;   077037ACFAAAE3BF5 CRC64;

Query Match            45.7%;  Score 878.5;  DB 5;  Length 472;
```

```
Best Local Similarity 42.6%; Pred. No. 2.8e-67;
Matches 181; Conservative 64; Mismatches 83; Indels 97; Gaps 6;

Qy   36 PGLDLQAIELAAQSNH------HCH-AQKGPDSHCDPKKGKAQRQLYVASAICLLEMIGE   88
           :|||:        ||  |  |:|:  |:|:||:|:|:|:||:|||||
Db   53 PATPAQIFCLHGRSNNVEVRDHCHRARSEGVDV------KARRKLIIASICLCLVFMIAE  105

Qy   89 VVGGYLAHSLAVMTDAAHLLTDFASMLISLFSLWMSSRPATKTMNFGWQRAEILGALVSV  148
        :||| |:|:||:|||||||||||||:|||||::|:|.||:|:|:|||:|||:||.|:||:
Db  106 IVGGVLSNSLAIATDAAHLLTDFASFMISLFAIWIAGRPSTQRMSFGWYRAEVIGAMASV  165

Qy  149 LSIWVVTWGVLYLYLAVERLISGDYEIDGGTMLITSGCAVAVNIIMGLLTLHQ-----SGH  201
        |.|||:|.|||:|:|||:||||||||:::||||||||.|:||||||||||||       |||
Db  166 FMIWVITGILVWLAIGRLISGDYEVANAKIMLITSGLAILVNVIMGVLQLQHGHSHGLGGGH  225

Qy  202 GHSHG--------TTNQQEENPS-------                                 216
        |||||        :|||||||:
Db  226 GHSHGGSKNASHVQATSTPCSDSPSQRIEGGVAYAPEDAELPGGGLPTFSYQNTKLVDPT  285

Qy  217 ------VRAAFIHVIGDFMQSMGVLVAAYILYFKP                           245
              |:|||:|||||||::||:.|||||:||:||:
Db  286 LDLEIAAVLAETAPGSHHHGGPVGREAVNMNVRAALIHVIGDVIQSVGVFVAAGVIYFWP  345

Qy  246 EYKYVDPICTFVFSILVLGTTILTILRDVILVILMEGTPKGVDFTAVRDLLSVEGVEALHS  305
        |::|:||||||||||:||.||:|:|::|.:|:||||||| ..:|:|||||||.|||||:|:
Db  346 EYSIVDPICTFVFSIIVLFTTFTIMKDALLVLMEGTPNYMIHVAEVLQIFQGIEGVERVHN  405

Qy  306 LHIWALTVAQPVLSVHIAIAQNTDAQAVLKTASSRLQGKFHFHTVTIQIEDYSEDMKDCQ  365
        |.|||:. ..| |:.||.||:|:. |.||.::::.  |:|||| :||||||||:::|::|
Db  406 LRIWALSINKVALSAHLAIAENANPKRILDAATSAVHLRYNFFETTIQIEDYTAQMESCL  465

Qy  366 ACQGP   370
          |.|
Db  466 QCNVP   470


RESULT 5
Q22541

ID   Q22541       PRELIMINARY;      PRT;      382 AA.
AC   Q22541;
DT   01-NOV-1996 (TREMBLrel. 01, Created)
DT   01-NOV-1996 (TREMBLrel. 01, Last sequence update)
DT   01-JUN-2002 (TREMBLrel. 21, Last annotation update)
DE   T18D3.3 protein.
GN   T18D3.3.
OS   Caenorhabditis elegans.
OC   Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
OC   Rhabditidae; Peloderinae; Caenorhabditis.
OX   NCBI_TaxID=6239;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   McMurray A.A.;
RL   Submitted (NOV-1995) to the EMBL/GenBank/DDBJ databases.
RN   [2]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=99069613; PubMed=9851916;
RA   none;
RT   "Genome sequence of the nematode C. elegans: A platform for
RT   investigating biology.";
RL   Science 282:2012-2018(1998).
DR   EMBL; Z68119; CAA92193.1;
DR   InterPro; IPR002524; Cation_effux.
DR   Pfam; PF01545; Cation_effux; 1.
DR   TIGRFAMs; TIGR01297; CDF; 1.
SQ   SEQUENCE 382 AA;  44A02AEF60B89484 CRC64;

Query Match 43.6%; Score 838.5; DB 5; Length 382;
Best Local Similarity 48.6%; Pred. No. 6e-64;
Matches 161; Conservative 66; Mismatches 93; Indels 11; Gaps 3;

Qy   50 NHHCHAQ-KGPDSHCDPKKGKAQRQLVASAICLLFMIGEVVGGYLAHSLAVMTDAAHLL  108
        | |||:  :||||||||||||||||| :||||||||::||| || |:|:||:||||||||
Db   53 NFHCHDEADSTDSH--DSNRRATRIIWLTVVLCLFMVCEVIGGVLAGSLAIVTDAAHLL  110
```

```
RESULT 6
Q9VKA3

ID   Q9VKA3       PRELIMINARY;      PRT;     1677 AA.
AC   Q9VKA3;
DT   01-MAY-2000 (TREMBLrel. 13, Created)
DT   01-MAY-2000 (TREMBLrel. 13, Last sequence update)
DT   01-JUN-2002 (TREMBLrel. 21, Last annotation update)
DE   CG17215 protein.
GN   CG17215.
OS   Drosophila melanogaster (Fruit fly).
OC   Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC   Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC   Ephydroidea; Drosophilidae; Drosophila.
OX   NCBI_TaxID=7227;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=BERKELEY;
RX   MEDLINE=20196006; PubMed=10731132;
RA   Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA   Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
RA   George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA   Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
RA   Brandon R.C., Rogers Y.-H.C., Blazej R.G., Champe M., Pfeiffer B.D.,
RA   Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
RA   Abril J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
RA   Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA   Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
RA   Borkova D., Botchan M.R., Bouck J., Brokstein P., Brottier P.,
RA   Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA   Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
RA   de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA   Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA   Durbin K.J., Evangelista C.C., Ferraz C., Ferriera S., Fleischmann W.,
RA   Fosler C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,
RA   Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA   Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
RA   Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
RA   Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA   Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA   Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA   Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
RA   Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
RA   Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA   Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacleb J.M.,
RA   Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA   Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
RA   Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
RA   Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
RA   Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA   Wang Z.-Y., Wassarman D.A., Weinstock G.M., Weissenbach J.,
RA   Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
```

```
Qy  109 TDFASMLISLFSLWMSSRPATKTMNFGWQRAEILGALVSVLSIWVVTWGVLYLYLAVERLIS  168
        |||||:|||||||::||:|:||:||||:||:|||.::.:.:::::|:||:|::::|||||:|
Db  111 TDFASVLISLFSLYIARRPPSQKMSFGFHRAEVLGAFFSVFLIWVTGVLVVLAIMRIVS    170

Qy  169 GDYEIDGGTMLITSGCAVAVNIIMGLTLHQSGHGHSHGTTNQ------QEENPSVRAA    220
        |||:::||:|.:|:..:.||||::|.|::...|:||:||      |||||||.||||
Db  171 GDYEVEGGIMALTAAILGVVNVLMLALLYFGGHSHSHGGGSSHSHSHGGGNGDNINVRAA  230

Qy  221 FIHVIGDFMQSMGVLVAAYILYFKPEYKYVDPICTFVFSILVLGTTILTILRDVILVLMEG  280
        |||::||::||:|||||| :::||:|.::::|||||:||::||.|||:.|||||:::::||
Db  231 FIHVLGDLLQSLGVLVAALFIYFQPSWVIIDPICTLVFSVILVCTTIYILRDAMIVLLEG   290

Qy  281 TPKGVDFTAVRDLLLSVEGVEALHSLHIWALTVAQPVLSVHIAIAQNTDAQAVLKTASSR  340
           .:|:||:.||:.||:.||::|:::|.||:||:|:.|||:| :    :.|||:::||
Db  291 RPSNIDFAKVFSSLEDIEGVKKVHDLRIWSLTMDKIALSVHLEIDANSQSQSILRETRKM  350

Qy  341 LQGKFHFHTVTIQIEDYSEDMKDCQACQGPS   371
        |:|..:.|.:||||||.::.:.||| | |
Db  351 LKQTYNVHEITIQIEEFGANRSDCGKCDPPT   381
```

```
RA   Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA   Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
RA   Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
RT   "The genome sequence of Drosophila melanogaster.";
RL   Science 287:2185-2195(2000).
DR   EMBL; AE003635; AAF53175.1; -.
DR   FlyBase; FBgn0032418; CG17215.
DR   InterPro; IPR000337; GPCR_Mgr.
DR   InterPro; IPR002524; Cation_efflux.
DR   InterPro; IPR000834; Zn_carbopept.
DR   Pfam; PF01545; Cation_efflux; 1.
DR   TIGRFAMs; TIGR01297; CDF; 1.
DR   PROSITE; PS00133; CARBOXYPEPT_ZN_2; UNKNOWN_1.
DR   PROSITE; PS00259; G_PROTEIN_RECEP_F3_4; 1.
SQ   SEQUENCE 1677 AA; 188028 MW; F50A8D282A8E3B6E CRC64;
```

```
Query Match     42.2%;  Score 811.5;  DB 5;  Length 1677;
Best Local Similarity   37.3%;
Matches  172;  Conservative  80;  Mismatches  114;  Indels  95;  Gaps  6;

Qy    4 KEKQHLLDARPAIRSYT------------------------GSLWQEGAGW 30
          |:||| :| : |                              ||||
Db  684 KNSENALDAKPQTATNTEETHGCINILKVPFQSNLIYYPEKPSRQDKGSGTSDFQPGAPE 743

Qy   31 IPLPRPGL---DLQAILEAAQS-----------------NHHCHAQKGPDSHCDEKK 67
           | || :    : |:|||                    ||:||||||| |   |
Db  744 TPIVENSAVDSDKRAVEIMPENVKNSEEKKIDNSDSTKTVTITGHSHITAKWDHCHEKE 803

Qy   68 GK-----AQRQLYVASAICLLFMIGEVVGGYLAHSLAVMTDAAHLLTDFASMLISLFSL 121
          |       ::|:||:||:|| ||||||:|||| ||:|:|||||||||:|| ||:| :|
Db  804 REIGVDKRAARRVLIIACILCTIFLILEVIGGILSNSLAIATDAAHLLTDLASFLISISAL 863

Qy  122 WMSSRPATKTMNFGWQRAEILGALVSVLSIWVVTGVLVLYLAVERLISGDYEIDGGTMLIT 181
          | : ||: : :| ||:||::||::||: |||||| |: | |:  :   : |  | ||||
Db  864 HLAGRPSSERLNYGWHRAEVIGAMVSIFFIWVVTGILVYMAIMRWVNQDFELDAKIMLIT 923

Qy  182 SGCAVAVNIIMGLTLHQSGHGHSHGTTN-----------------QQ 211
           . | ||||:|:  |:|||:|||                      ||
Db  924 SALAILFNVIMAMQLQ---HGHSHSLPGVHKMSKDAGSVLGSRMILLGCKSVSMQYAARG 980

Qy  212 EENPSVRAAFIHVIGDFMQSMGVLVAAVILYFKPEYKIVDPICTFVFSILVLGTTLTILR 271
          || :| |||:||::||::|| ||::||:|::| ||| : |  |||||| :||::| | ||
Db  981 HENINVRAAIIHVVGDIIQSFGVFVAALIIFFWPEWAEFMDSVCTFVFSVLVLVVTEKILR 1040

Qy  272 DVILVLMEGTPKGVDFTAVRDLLLSVEGVEALHSLHIWALTVAQPVLSVHIAIAQNTDAQ 331
          || ::||: ||  :|: :|: :||:| ||| :||::|||:  :|| ||:|:::|: | |
Db 1041 DVLMLMEATPDFMDYEEVKQTFLSISGVEHVHNLRIWALSINKVALSAHLAISKDADPQ 1100

Qy  332 AVLKTASSRLQGSKFHFHVTIQIEDYSEDMKDCQACQGPSD 372
           :|  : : | :.|| |: ||||::| ||::||:||  |||
Db 1101 LILEEATTLIHKRFKFFETTIQIEEYSPGMENCGQCLSPSD 1141
```

```
RESULT 7
ID   Q8TC39        PRELIMINARY;    PRT;   429 AA.
AC   Q8TC39;
DT   01-JUN-2002 (TrEMBLrel. 21, Created)
DT   01-JUN-2002 (TrEMBLrel. 21, Last sequence update)
DT   01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE   Solute carrier family 30 (zinc transporter), member 4.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=TESTIS;
RA   Strausberg R.;
RL   Submitted (MAR-2002) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; BC026089; AAH26089.1; -.
SQ   SEQUENCE 429 AA; 47482 MW; 97B7FCBE881C8C32 CRC64;
```

```
Query Match     37.6%;  Score 722.5;  DB 4;  Length 429;
Best Local Similarity   39.0%;
Matches  151;  Conservative  88;  Mismatches  117;  Indels  31;  Gaps  7;

Qy    2 EAKEKQHLLDARPAIRSYTGSLWQEGAGWIPLPRPGLDLQAIELAAQSNHHCHAQKGPDS 61
          ||  :  ||:  :  ::      :         |  | :|||||||:|:|||||:|:|:
Db   56 EAPERP-VNGAHPTLQADDDSLLDQD--LPLTNSQLSLK-----VDSCDNCSKQRE--- 103

Qy   62 HCDPKKGKAQRQLYVASAICLLFMIGEVVGGYLAHSLAVMTDAAHLLTDFASMLISLFSL 121
               : :|:|:|| | |:||||||||:||||||| |||:|||||:|:||:||:| | |
Db  104 --ILKQRKVKARLTIAAVLLFMIGELVGGYIANSLAIMTDALHMLTDLSAIILTLAL 161

Qy  122 WMSSRPATKTMNFGWQRAEILGALVSVLSIWVVTGVLVLYLAVERLISGDYEIDGGTMLIT 181
          | ||:  ||:  ||:||:|:|:|:||:| ::::|:|:| | ||:::  :  |: ||:|||
Db  162 WLSSKSPTKRFTFGFHRLEVLSAMISVLLVVIILMGFLLYEAVQRTIHMNYEINGDILMLIT 221

Qy  182 SGCAVAVNIIMGLTLHQSGHGHSHGTT------------------NQQEENPSVRAAFIHVIGD 227
           . || ||:|||  |:|||:|:||          :         | |::| |||||::|:||
Db  222 AAVGVAVNVIMGFLLNQSGHRHSHSHSLPSNSPTRGSGCERNHGQDSLAVRAAFVHALGD 281

Qy  228 FMQSMGVLVAAYILYFKPEYKYVDPICTFVFSILVLGTTLTILLRDVILVLMEGTPKGVDF 287
          ::|| ||:|||:|: ||||| : ||||| ||| ||  ||: | | |:|:::||| |  :|
Db  282 LVQSVGVLIAAYIIRFKPEYKIADPICTYVFSLLVAFTTFRIIWDTVVIILEGVPSHLNV 341

Qy  288 TAVRDLLLSVEGVEALHSLHIWALIVAQPVLSVHIAIAQNTDA--QAVLKTASSRLQGKF 345
           | ::|: :: :||:| : :::: ::: |  :||: : | :    : |  : :     ::
Db  342 DYIKEALMKIEDVYSVEDLNIWSLISGGKSTAIVHIQLIPGSSSSKWEEVQSKANHLLNTF 401

Qy  346 HFHTVTIQIEDYSEDM-KDCQACQGPS 371
          . : | ||::|   :: : |: || |
Db  402 GMYRCTIQLQSYRQEVDRTCANCQSSS 428
```

```
RESULT 8
ID   O45923        PRELIMINARY;    PRT;   410 AA.
AC   O45923;
DT   01-JUN-1998 (TrEMBLrel. 06, Created)
DT   01-JUN-1998 (TrEMBLrel. 06, Last sequence update)
DT   01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE   Y39E4A.2b protein.
GN   Y39E4A.2B.
OS   Caenorhabditis elegans.
OC   Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
OC   Rhabditidae; Peloderinae; Caenorhabditis.
OX   NCBI_TaxID=6239;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Barlow K.;
RL   Submitted (JAN-1998) to the EMBL/GenBank/DDBJ databases.
RN   [2]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=99069613; PubMed=9851916;
RA   none;
RT   "Genome sequence of the nematode C.elegans: A platform for
RT   investigating biology.";
RL   Science 282:2012-2018(1998).
DR   EMBL; AL021480; CAA16328.1; -.
DR   InterPro; IPR002524; Cation_efflux.
DR   Pfam; PF01545; Cation_efflux; 1.
DR   TIGRFAMs; TIGR01297; CDF; 1.
SQ   SEQUENCE 410 AA; 44082 MW; 8379EA94I809270C CRC64;
```

```
Query Match     36.9%;  Score 709.5;  DB 5;  Length 410;
Best Local Similarity   40.8%;
Matches  139;  Conservative  73;  Mismatches  104;  Indels  25;  Gaps  4;

Qy   52 HCHAQKGPDSHCD-------------------PKKGKAQRQLYVASAICLLFMIGEVVGGYL 94
          | : : |  : |                    ||| |||||||||||:|||||||:||| |
Db   66 HSHTHNNDDSSSDCSGAGGGAHKHSHDEKYQKGRRAEKVLWAVAALSAVFIAAEVGGFW 125

Qy   95 AHSLAVMTDAAHLLTDFASMLISLFSLWMSSRPATKTMNFGWQRAEILGALVSVLSIWVV 154
          :||||||||||||||| |:|| :|: |:|:::||| : :||::|: ::::::|::|:|::
```

```
Db  126 AQSLAIMTDAGHHMLSDLLSFIISFAIRCARLPASKRLSFGYERAEVLGALTSVIILWVL 185

Qy  155 TGVLVYLAVERLISGDYEIDGGTMLITSGGAVAVNIIMGLTLH--QSGHGHSHGTTN--- 209

Qy  186 TFVLVVVAIQRIVNNEHVDADVMLITAGVGVLFNIVMGLVLHFGTGGHGTHGGHSSHG 245

Qy  210 --QQEENPSVRAAFIHVIGDFMQSMGVLVAAYILYFKPEYKYVDPICTFVFSILVLGTTL 267

Db  246 HAHDGKNVNVRAALIHVIGDLVQSIGVLIIAALIIRF-TGWTLADPICTFLFSIIVLFTTV 304

Qy  268 TIIRDVILVLMEGTPKGVDFTAVRDLLLSVEGVEALHSLHIWALTVAQPVLSVHIAIAQN 327

Db  305 TVMRDIFFVLMEATPSHYDLSDVKKALSALEGVKGVHDLHLWSIGMDKTAFSVHLALESP 364

Qy  328 TDAQAVLKTASSRLQGKFHFHTVTIQIEDYSEDMKDCQACQ 368

Db  365 NRAMENVAEARSLIRRRFGVAVATVQVEPFDEKIDSCDTCQ 405
```

```
RESULT 9

ID   O45922     PRELIMINARY;     PRT;    391 AA.
AC   O45922;
DT   01-JUN-1998 (TrEMBLrel. 06, Created)
DT   01-JUN-1998 (TrEMBLrel. 06, Last sequence update)
DT   01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE   Y39E4A.2a protein.
DE   Y39E4A.2a
GN   none;
OS   Caenorhabditis elegans.
OC   Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
OC   Rhabditidae; Peloderinae; Caenorhabditis.
OX   NCBI_TaxID=6239;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Barlow K.;
RL   Submitted (JAN-1998) to the EMBL/GenBank/DDBJ databases.
RN   [2]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=99069613; PubMed=9851916;
RA   "Genome sequence of the nematode C.elegans: A platform for
RT   investigating biology.";
RL   Science 282:2012-2018(1998).
DR   EMBL; AL021480; CAA16327.1; -.
DR   InterPro; IPR002524; Cation_efflux.
DR   Pfam; PF01545; Cation_efflux; 1.
DR   TIGRFAMs; TIGR01297; CDF; 1.
SQ   SEQUENCE 391 AA; 42389 MW; 41F1260CEA4F6BD6 CRC64;

Query Match   36.5%;   Score 702; DB 5;   Length 391;
Best Local Similarity   40.6%;   Pred. No. 3.4e-52;
Matches 138; Conservative 76; Mismatches 102; Indels 24; Gaps 5;
```

```
Qy   52 HCH--AQKGPDSHCDPKKG--------------KAQRQLYVASAICLLFMIGEVVGGYLA 95

Db   48 HCHYWDENDDDMYARVERGGSTDSASSREDTGRRAEKVLWAVAALSAVFIAAEFVGGFWA 107

Qy   96 HSLAVMTDAAHLLTDFASMLISLFSLWMSSRPATKTMNFGWQRAEILGALVSVLSIWVVT 155

Db  108 QSLAIMTDAGHMLSDLLSFIISFAIRCARLPASKRLSFGYERAEVLGALTSVIILWVLT 167

Qy  156 GVLVYLAVERLISGDYEIDGGTMLITSGGAVAVNIIMGLTLH---QSGHGHSHGTTN--- 209

Db  168 TVLVVVAIQRIVNNEHVDADVMLITAGVGVLFNIVMGLVLHFGTGGHGTHGGHSSHGH 227

Qy  210 -QQEENPSVRAAFIHVIGDFMQSMGVLVAAYILYFKPEYKYVDPICTFVFSILVLGTTL 268

Db  228 AHDGKNVNVRAALIHVIGDLVQSIGVLIIAALIIRF-TGWTLADPICTFLFSIIVLFTTV 286

Qy  269 ILRDVILVLMEGTPKGVDFTAVRDLLLSVEGVEALHSLHIWALTVAQPVLSVHIAIAQNT 328

Db  287 VMRDIFFVLMEATPSHYDLSDVKKALSALEGVKGVHDLHLWSIGMDKTAFSVHLALESPN 346
```

```
Qy  329 DAQAVLKTASSRLQGKFHFHTVTIQIEDYSEDMKDCQACQ 368

Db  347 RAMENVAEARSLIRRRFGVAVATVQVEPFDEKIDSCDTCQ 386
```

```
RESULT 10

ID   081036     PRELIMINARY;     PRT;    398 AA.
AC   081036;
DT   01-NOV-1998 (TrEMBLrel. 08, Created)
DT   01-NOV-1998 (TrEMBLrel. 08, Last sequence update)
DT   01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE   Zinc transporter (ZAT).
GN   AT2G46800.
OS   Arabidopsis thaliana (Mouse-ear cress).
OC   Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC   Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC   eurosids II; Brassicales; Brassicaceae; Arabidopsis.
OX   NCBI_TaxID=3702;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=cv. COLUMBIA;
RX   MEDLINE=20083487; PubMed=10617197;
RA   Lin X., Kaul S., Rounsley S.D., Shea T.P., Benito M.-I., Town C.D.,
RA   Fujii C.Y., Mason T.M., Bowman C.L., Barnstead M.E., Feldblyum T.V.,
RA   Buell C.R., Ketchum K.A., Lee J.J., Ronning C.M., Koo H., Moffat K.S.,
RA   Cronin L.A., Shen M., Vanaken S.E., Umayam L., Tallon L.J., Gill J.E.,
RA   Adams M.D., Carrera A.J., Creasy T.H., Goodman H.M., Somerville C.R.,
RA   Copenhaver G.P., Preuss D., Nierman W.C., White O., Eisen J.A.,
RA   Salzberg S.L., Fraser C.M., Venter J.C.;
RT   "Sequence and analysis of chromosome 2 of the plant Arabidopsis
RT   thaliana.";
RL   Nature 402:761-768(1999).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=cv. COLUMBIA;
RA   Lin X.;
RL   Submitted (MAR-2000) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AC005310; AAC33498.1; -.
DR   InterPro; IPR002524; Cation_efflux.
DR   Pfam; PF01545; Cation_efflux; 1.
DR   TIGRFAMs; TIGR01297; CDF; 1.
SQ   SEQUENCE 398 AA; 43827 MW; 7E20E0B292237BB23 CRC64;

Query Match   33.8%;   Score 649.5;   DB 10;   Length 398;
Best Local Similarity   37.9%;   Pred. No. 1.1e-47;
Matches 134; Conservative 71; Mismatches 88; Indels 61; Gaps 4;
```

```
Qy   60 DSHCDPKKGKRAQRQLYVASAICLLFMIGEVVGGYLAHSLAVMTDAAHLLTDFASMLISLF 119

Db   47 DAH---ERSASMRKLCIAVVLCLVFMSVEVVGGIKANSLAILTDAAHLLSDVAAFAISLF 103

Qy  120 SLWMSSRPATKTMNFGWQRAEILGALVSVLSIWVVTGVLVYLAVERLISGDYEIDGGTML 179

Db  104 SLWAAGWEATPRQTGFFRIEILGALVSIQLIWLLTGILVVEAIIRIVIVTETSEVNGFLMF 163

Qy  180 ITSGCAVAVNIIMGLTL----HQSGHGHSHG----------------TTN---------- 209

Db  164 LVAAFGLVVNIIMAVLLGHDHGHSHGHDHHNHSHGVTVTTHHHHHHGHGHSHG 223

Qy  210 ---------------------QQEENPSVRAAFIHVIGDFMQSMGVLVAAYIL 241

Db  224 HGEDKHHAHGDVTEQLLDKSKTQVAAKEKRKRNINLQGAYLHVLGDSIQSVGVMIGGAII 283

Qy  242 YFKPEYKYVDPICTFVFSILVLGTTLILRDVILVLMEGTPKGVDFTAVRDLLLSVEGVE 301

Db  284 WYNPEWKIVDLICTLAFSVIVLGTTIINMILRNILEVLMESTPREIDATKLEKGLLEMEEVV 343

Qy  302 ALHSLHIWALTVAQPVLSVHIAIAQNTDAQAVLKTASSRLQGKFHFHTVTIQIE 355

Db  344 AVHELHIWAITVGKVLLACHVNIRPEADADMVLNKVIDYIRREYNISHVTIQIE 397
```

RESULT 11
Q94A29
ID   Q94A29   PRELIMINARY;   PRT;   360 AA.
AC   Q94A29;
DT   01-DEC-2001 (TrEMBLrel. 19, Created)
DT   01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT   01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE   Putative vacuolar metal-ion transport protein MTP1t2.
GN   MTP1.
OS   Thlaspi goesingense.
OC   Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC   Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC   eurosids II; Brassicales; Brassicaceae; Thlaspi.
OX   NCBI_TaxID=60913;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=21396513; PubMed=11481436;
RA   Persans M.W., Nieman K., Salt D.E.;
RT   "Functional activity and role of cation-efflux family members in Ni
RT   hyperaccumulation in Thlaspi goesingense.";
RL   Proc. Natl. Acad. Sci. U.S.A. 98:9995-10000(2001).
DR   EMBL; AY044454; AAK91871.1; -.
DR   InterPro; IPR002524; Cation_efflux.
DR   Pfam; PF01545; Cation_efflux; 1.
DR   TIGRFAMs; TIGR01297; CDF; 1.
SQ   SEQUENCE   360 AA; 39555 MW; 612BAFAE8DB2301F CRC64;

Query Match   33.7%; Score 649; DB 10; Length 360;
Best Local Similarity   42.4%; Pred. No. 1.1e-47;
  Matches 133; Conservative   65; Mismatches   92; Indels   24; Gaps   3;

Qy   66 KKGKAQRQLYVASAICLLFMIGEVVGGYLAHSLAVMTDAAHLLTDFASMLISLFSLWMSS 125
          ::| :|| | | ||||| ||||| ||||: ||||||||||||| |:|||||||
Db   46 ERNASMRKLCIAVVLCLVFMSVEVVGGVKANSLAIMTDAAHLLSDVAAFAISLFSLWAAG 105

Qy   126 RPATKTMNFGWQRAEILGALVSVLSIWVVTGVLVYLAVERLISGDYEIDGGTMLITSGCA 185
          |:| ||||||| || ||| | || |||||||||| |: : ||| |
Db   106 WEATPRHYGFFRIELGALVSVIQLIWILTGILVYEAIIRLLTETSEVNGFLMFAVATFG 165

Qy   186 VAVNIIMGLTL---------HQSGHGHSHG---------------TTNQQEE----NPSVRAAF 221
          ::|||| ||| | |||||||
Db   166 LVNIIMAVMLGHDHDNGHGHSHGEDKHDEAHGDVTEQLLEKPKQQKEKKKRNINAQGAY 225

Qy   222 IHVIGDEMQSMGVLVAAYILYFKPEYKVVDPICTFVFSILVLGTTLTLIRDVILVLMEGT 281
          :||:|| ::: ||||:| :|: ||: |||: ||:||| ||||: ||| ||||: ||||
Db   226 LHVLGDSIOSVGVMIGGAAAIWNPKWKIIDLICTLVFSVIVLGTTINMIRSILEVLMEST 285

Qy   282 PKGVDFTAVRDLLLSVEGVEALHSLHIWALTVAQPVLSVHIAIAQNTDAQAVLKTASSRL 341
          |: || || :: : | | : ||||||| || | :|: || | |
Db   286 PREIDATKLEKGLLEMEEVVAVHELHIWASTVGKVLLACHVNARPDADADMVLNKVVDYI 345

Qy   342 QGKFHFHTVTIQIE 355
          : |||||
Db   346 RREYNISHVTIQIE 359

RESULT 12
Q92T63
ID   Q92T63   PRELIMINARY;   PRT;   398 AA.
AC   Q92T63;
DT   01-MAY-1999 (TrEMBLrel. 10, Created)
DT   01-MAY-1999 (TrEMBLrel. 10, Last sequence update)
DT   01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE   Zinc transporter ZAT.
OS   Arabidopsis thaliana (Mouse-ear cress).
OC   Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC   Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC   eurosids II; Brassicales; Brassicaceae; Arabidopsis.
OX   NCBI_TaxID=3702;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=99169295; PubMed=10069843;
RA   van der Zaal B.J., Neuteboom L.W., Pinas J.E., Chardonnens A.N.,
RA   Schat H., Verkleij J.A.C., Hooykaas P.J.J.;
RT   "Overexpression of a novel Arabidopsis gene related to putative zinc-
RT   transporter genes from animals can lead to enhanced zinc resistance
RT   and accumulation.";
RL   Plant Physiol. 119:1047-1055(1999).
DR   EMBL; AF072858; AAD11757.1; -.
DR   InterPro; IPR002524; Cation_efflux.
DR   Pfam; PF01545; Cation_efflux; 1.
DR   TIGRFAMs; TIGR01297; CDF; 1.
SQ   SEQUENCE   398 AA; 44063 MW; 6FC33112235D7C3C CRC64;

Query Match   33.5%; Score 643.5; DB 10; Length 398;
Best Local Similarity   37.0%; Pred. No. 3.7e-47;
  Matches 131; Conservative   71; Mismatches   91; Indels   61; Gaps   3;

Qy   60 DSHCDPKKGKAQRQLYVASAICLLFMIGEVVGGYLAHSLAVMTDAAHLLTDFASMLISLF 119
          | | :| :|| | | ||||| ||||| ||||: ||||||||||||| |:|||
Db   47 DAH--ERSASMRKLCIAVVLCLVFMSVEVVGGVKANSLAIMTDAAHLLSDVAAFAISLF 103

Qy   120 SLWMSSRPATKTMNFGWQRAEILGALVSVLSIWVVTGVLVYLAVERLISGDYEIDGGTML 179
          ||| | | |:| ||||||| || ||| | || |||||||||| |: :||
Db   104 SLWAAGWEATPRQTYGFFRIELGALVSVIQLIWILTGILVYEAIIRLVTETSEVNGFLMF 163

Qy   180 ITSGCAVAVNIIMGLTL--------------------HQSGHGHS 204
          | ::|||| |||
Db   164 LVAAFGLVVNIIMAVLLGHDHGHGPQSWTWAWHGHDHNHSHGVTVTHHHHHDHEHGHSHG 223

Qy   205 HG-------------TTNQQEENPSVRAAFIHVIGDFMQSMGVLVAAYIL 241
          || ||::: |||||: |:||||| |: :||
Db   224 HGEDKHHAHGDVTEQLLDKSKTQVAAKEKRKRNINLQGAYLHVLGDSIOSVGVMIGGAII 283

Qy   242 YFKPEYKVVDPICTFVFSILVLGTTLTLIRDVILVLMEGTPKGVDFTAVRDLLLSVEGVE 301
          |: ||: |||: ||:||| ||||: ||| ||||: |||| |: || || :: : | |
Db   284 WYNPEWKIVDLICTLAFSVIVLGTTINMIRNILEVLMESTPREIDATKLEKGLLEMEEVV 343

Qy   302 ALHSLHIWALTVAQPVLSVHIAIAQNTDAQAVLKTASSRLQGKFHFHTVTIQIE 355
          : ||||||| || | :|: || | | : |||||
Db   344 AVHELHIWAITVGKVLLACHVNIRPEADADMVLNKVIDYIRREYNISHVTIQIE 397

RESULT 13
Q9LXS1
ID   Q9LXS1   PRELIMINARY;   PRT;   378 AA.
AC   Q9LXS1;
DT   01-OCT-2000 (TrEMBLrel. 15, Created)
DT   01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT   01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE   Zinc transporter-like protein.
GN   T20N10.160.
OS   Arabidopsis thaliana (Mouse-ear cress).
OC   Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC   Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC   eurosids II; Brassicales; Brassicaceae; Arabidopsis.
OX   NCBI_TaxID=3702;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   D'Angelo M., Vezzi A., Modesto D., Pigazzi M., Valle G., Mewes H.W.,
RA   Rudd S., Lemcke K., Mayer K.F.X., Quetier F., Salanoubat M.;
RL   Submitted (APR-2000) to the EMBL/GenBank/DDBJ databases.
RN   [2]
RP   SEQUENCE FROM N.A.
RL   Submitted (APR-2000) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AL353032; CAB88298.1; -.
DR   InterPro; IPR002524; Cation_efflux.
DR   Pfam; PF01545; Cation_efflux; 1.
DR   TIGRFAMs; TIGR01297; CDF; 1.
SQ   SEQUENCE   378 AA; 41221 MW; 7C05E7D542B69157 CRC64;

Query Match   33.4%; Score 642; DB 10; Length 378;
Best Local Similarity   39.3%; Pred. No. 4.7e-47;
  Matches 128; Conservative   69; Mismatches   93; Indels   36; Gaps   2;

```
Qy      66 KKGKAQRQLYVASAICLLFMIGEVVGGYLAHSLAVMTDAAHLLTDFASMLISLFSLWMSS 125
           ::|||:||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      51 ERNASMRKLIAVLCAIFIVVEVVGGIKANSLAILTDAAHLLSDVAAFAISLFSLWASG 110

Qy     126 RPATKTMNFGWQRAEILGALVSVLSLSIWVVTVGVLVLAVERLISGDYEIDGGTMLITSGCA 185
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     111 WKANPQQSYGFFRIEILGALVSIQMTWLLAGTLVYEAIVRLNNGSGEVEGSLMFAVSAVG 170

Qy     186 VAVNIIMGLTL--------------------HQSGHGHSHG-------------------- 209
           |:|||:|||                      ::|||||||
Db     171 LLVNIAMAILLGHDGHGHSHDNGHGHSHDHGHGIAATEHHHDSGHDESQLSDVLIEQ 230

Qy     210 QQEENPSVRAAFIHVIGDFMQSMGVLVAAYILYFKPEYKYVDPICTFVFSILVLGTTLTI 269
           |||||||:|||||||||||:|:||||||||||||||||| ||||||||||:|||||||:
Db     231 KKQRNVNIQGAYLHVLGDSIQSVGVMIGGAIIWYKPEWKIDLICTLVFSVIVLGTTIGM 290

Qy     270 LRDVILVLMEGTPKGVDFTAVRDLLLSVEGVEALHSLHIWALTVAQPVLSVHIAIAQNTD 329
           |||:|:||||||||||||||||||||||||||||||||||||||:||:||||:|:|:||
Db     291 LRNILEVLMESTPREIDPTMLEKGVCEIEEVVAVHELHIWAITVGKLLLACHVKIRPEAE 350

Qy     330 AQAVLKTASSRLQGKFHFHIVTIQIE 355
           ||||||||||||||||||||||||||
Db     351 ADMVLDKIIDYIKREHNISHVTIQIE 376

RESULT 14
Q93XE9
ID  Q93XE9        PRELIMINARY;    PRT;    396 AA.
AC  Q93XE9;
DT  01-DEC-2001 (TrEMBLrel. 19, Created)
DT  01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT  01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE  Zinc transporter.
GN  ZTP1.
OS  Thlaspi caerulescens.
OC  Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC  Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC  eurosids II; Brassicales; Brassicaceae; Thlaspi.
OX  NCBI_TaxID=107243;
RN  [1]
RP  SEQUENCE FROM N.A.
RC  TISSUE=ROOT;
RA  Assuncao A.G.L., da Costa Martins P., de Folter S., Vooijs R.,
RA  Schat H., Aarts M.G.M.;
RT  "Elevated expression of metal transporter genes in three accessions of
RT  the metal hyperaccumulator Thlaspi caerulescens.";
RL  Submitted (JUN-2000) to the EMBL/GenBank/DDBJ databases.
DR  EMBL; AF275750; AAK69428.1; -.
DR  InterPro; IPR002524; Cation_efflux.
DR  Pfam; PF01545; Cation_efflux; 1.
DR  TIGRFAMS; TIGR01297; CDF; 1.
SQ  SEQUENCE    396 AA;    43614 MW;    AD5224234302606085 CRC64;

Query Match             33.2%;    Score 639;    DB 10;    Length 396;
Best Local Similarity   36.6%;    Pred. No. 9.1e-47;
Matches  128;    Conservative  69;    Mismatches  93;    Indels  60;    Gaps  2;

Qy      66 KKGKAQRQLYVASAICLLFMIGEVVGGYLAHSLAVMTDAAHLLTDFASMLISLFSLWMSS 125
           ::|||:||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      46 ERNASMRKLIAVVCLVFMSVEIVGGIKANSLAIMTGAAHLLSDVAAFAISLFALWAAG 105

Qy     126 RPATKTMNFGWQRAEILGALVSVLSLSIWVVTVGVLVLAVERLISGDYEIDGGTMLITSGCA 185
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     106 WEATPROTYGFFRIEILGALVSIQLIWLLITGILVYEAIIRLLTETSEVNGFLMFAVATFG 165

Qy     186 VAVNIIMGLTL--------------------HQSGHGHSHG 206
           |:|||:|||                      ::|||||||
Db     166 LLVNIIMAVMLGHDHGHSHGHDHENHSGVTVTHDHDPDGHGHSHG 225

Qy     207 TTNQOE--------------------ENPSVRAAFIHVIGDFMQSMGVLVAAYILYFKP 245
           ::|||                     |::|:||||||||||:|:|||||||||||||||
Db     226 EDNQDEAHGDVTEQLLEKPKQEKEKKRNINLQGAYLHVLGDSIQSVGVMIGGAAIWYNP 285
```

```
Qy     246 EYKYVDPICTFVFSILVLGTTLTILRDVILVLMEGTPKGVDFTAVRDLLLSVEGVEALHS 305
           |||||||||||||||||||||:|:||||:|:|||||||||||||||||||||||||||||
Db     286 KWKIIDLICTLAFSVIVLGTTIGMLRNILEVLMESTPREIDATKLEKGLLEMEEVVAVHIE 345

Qy     306 LHIWALTVAQPVLSVHIAIAQNTDAQAVLKTASSRLQGKFHFHIVTIQIE 355
           ||||||||||||||||||||||||||||||||||||||||||||||||||
Db     346 LHIWAITVGKVLLACHVNVTPQADADMVLNKVVDYIRREYNISHVTVQIE 395

RESULT 15
Q94B01
ID  Q94B01        PRELIMINARY;    PRT;    392 AA.
AC  Q94B01;
DT  01-DEC-2001 (TrEMBLrel. 19, Created)
DT  01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT  01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE  Putative vacuolar metal-ion transport protein MTP1.
GN  MTP1.
OS  Thlaspi goesingense.
OC  Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC  Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC  eurosids II; Brassicales; Brassicaceae; Thlaspi.
OX  NCBI_TaxID=60913;
RN  [1]
RP  SEQUENCE FROM N.A.
RX  MEDLINE=21396513; PubMed=11481436;
RA  Persans M.W., Nieman K., Salt D.E.;
RT  "Functional activity and role of cation-efflux family members in Ni
RT  hyperaccumulation in Thlaspi goesingense.";
RL  Proc. Natl. Acad. Sci. U.S.A. 98:9995-10000(2001).
DR  EMBL; AY044452; AAK91869.1; -.
DR  InterPro; IPR002524; Cation_efflux.
DR  Pfam; PF01545; Cation_efflux; 1.
DR  TIGRFAMS; TIGR01297; CDF; 1.
SQ  SEQUENCE    392 AA;    43121 MW;    B1BADD63E045B490 CRC64;

Query Match             33.2%;    Score 638;    DB 10;    Length 392;
Best Local Similarity   38.4%;    Pred. No. 1.1e-46;
Matches  133;    Conservative  65;    Mismatches  92;    Indels  56;    Gaps  3;

Qy      66 KKGKAQRQLYVASAICLLFMIGEVVGGYLAHSLAVMTDAAHLLTDFASMLISLFSLWMSS 125
           ::|||:||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      46 ERNASMRKLCIAVVLCLVFMSVEVVGGVKANSLAIMTDAAHLLSDVAAFAISLFSLWAAG 105

Qy     126 RPATKTMNFGWQRAEILGALVSVLSLSIWVVTVGVLVLAVERLISGDYEIDGGTMLITSGCA 185
           |:|:|||||||||||||||::|:|||||:|||:|||||||||||||||||||||||||||
Db     106 WEATPROTYGFFRIEILGALVSIQLIWLLTGILVYEAIIRLLTETSEVNGFLMFAVATFG 165

Qy     186 VAVNIIMGLTL--------------------HQSGHGHSHG 206
           |:|||:|||                      ::|||||||
Db     166 LLVNIIMAVMLGHDHGHSHGHDHGNHSHDVTVTHDHDHDHDHDHDDGHGHSHGEDKH 225

Qy     207 --------------------TTNQOEE--------------------NPSVRAAFIHVIGDFMQSMGVLVAAYILYFKPEYKY 249
                               ::|||:                     |::|:|||||||||||:|:||||||||||||||
Db     226 DEAHGDVTEQLLEKPKQOKETKRNINVQGAYLHVLGDSIQGVGVMIGGAAIWYNPKKWI 285

Qy     250 VDPICTFVFSILVLGTTLTILRDVILVLMEGTPKGVDFTAVRDLLLSVEGVEALHSLHIW 309
           ||||||:|||||:||||:|:|||||:|:|||||||||||||||||||||||||||||||||
Db     286 IDLICTLVFSVIVLGTTINMIRSTLEVLMESTPREIDATKLEKGLLEMEEVVAVHELHIW 345

Qy     310 ALTVAQPVLSVHIAIAQNTDAQAVLKTASSRLQGKFHFHIVTIQIE 355
           ||:|||||||||||||||||||||||||||||||||||||||||||
Db     346 AITVGKVLLACHVNARPDADADMVLNKVVDYIRREYNISHVTIQIE 391

RESULT 16
Q94B00
ID  Q94B00        PRELIMINARY;    PRT;    392 AA.
AC  Q94B00;
DT  01-DEC-2001 (TrEMBLrel. 19, Created)
DT  01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT  01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
```

```
DE   Putative vacuolar metal-ion transport protein MTP1t1.
GN   MTP1.
OS   Thlaspi goesingense.
OC   Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC   Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC   eurosids II; Brassicales; Brassicaceae; Thlaspi.
OX   NCBI_TaxID=60913;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=21396513; PubMed=11481436;
RA   Persans M.W., Nieman K., Salt D.E.;
RT   "Functional activity and role of cation-efflux family members in Ni
RT   hyperaccumulation in Thlaspi goesingense.";
RL   Proc. Natl. Acad. Sci. U.S.A. 98:9995-10000(2001).
DR   EMBL; AY044453; AAK91870.1; -.
DR   InterPro; IPR002524; Cation_efflux.
DR   Pfam; PF01545; Cation_efflux; 1.
DR   TIGRFAMs; TIGR01297; CDF; 1.
SQ   SEQUENCE 392 AA; 43080 MW; D6C6910I7BD9194F CRC64;

Query Match             32.9%;  Score 633; DB 10; Length 392;
Best Local Similarity   38.4%;  Pred. No. 2.9e-46;
Matches  133;  Conservative  64;  Mismatches  93;  Indels  56;  Gaps  3;

Qy    66 KKGKAQRQLYASAICLLFMIGEVVGGYLAHSLAVMTDAAHLLTDFASMLISLFSLWMSS 125
          :: ||||:||||:|||: |||||||||||||| :||||||||||:||||||||| ||||
Db    46 ERNASMRKICIAVVLCLVFMSVEVVGVGVKANSLAIMTDAAHLLSDVAAFAISLFSLWAAG 105

Qy   126 RPATKIMNFGWQRAEILGALVSVLSIWVVTGVLVYLAVERLISGDYEIDGGTMLITSGCA 185
          :||| ||||||||||||||:|| :|||:||:|||| |: |: | :| |:|| ||| |
Db   106 WEATPROTYGFFRIEILGALVSIQLIWLLTGILVYEAIIRLPTETSEVNGFLMFAVAFFG 165

Qy   186 VAVNIIMGLTL------------------------HOSGHGHSH---------- 206
          :|||||| : |                         |||||||
Db   166 LIVNIIMAVMLGHDHGHSHGHGHGHGHGHSHDVTVTHDHDHDHDDGHGHSHGEDKH 225

Qy   207 -------------TTNQOEE---NPSVRAAFIHVIGDFMOSMGVLIVAAYILYFKPEYKY 249
                       ||:|::    ||| |  :|| :|  :| |:: ||  ||:|||
Db   226 DEAHGDVTEQLLEEPKQOKEKKKRNINVQGAVLHALGDSIOSVGVMICGAAIWYNPKWKI 285

Qy   250 VDPICFTVFSILVLGTTFILRDVILVLMEGTPKGVDFTAVRDLLLSVEGVEALHSLHIW 309
          :| || :|| |:||||| :|:|| ||||| ||   | |  : |||::| :| ||||||
Db   286 IDLICTLVFESVIVLGTTINMIRSILEVLMESTPREIDATKLEKGLLEMEEVVAVHELHIW 345

Qy   310 ALTVAQPVLSVHIAIAQNTDAQAVLKTASSRLGQKFHFHTVTIQIE 355
          |:||.: :|| |: :|| ||:|:|: :|: |  .:|||||:||||
Db   346 AITVGKVLLACHVNARPDADADMVLSKVVDYIRREYNISHVTIQIE 391
```

```
RESULT 17
ID   Q947R8        PRELIMINARY;     PRT;    421 AA.
AC   Q947R8;
DT   01-DEC-2001 (TrEMBLrel. 19, Created)
DT   01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT   01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE   Zinc transporter.
GN   EG2NT1.
OS   Eucalyptus grandis.
OC   Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC   Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC   eurosids II; Myrtales; Myrtaceae; Eucalyptus.
OX   NCBI_TaxID=71139;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=ROOT;
RA   Fairbairn D.J., Day S., Gomez-Gallego S., Sawbridge T., Teasdale R.D.;
RL   Submitted (OCT-1999) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AF197329; AAL25646.1; -.
DR   InterPro; IPR002524; Cation_efflux.
DR   Pfam; PF01545; Cation_efflux; 1.
DR   TIGRFAMs; TIGR01297; CDF; 1.
SQ   SEQUENCE 421 AA; 46695 MW; A42AA9AEFA2DA46A CRC64;

Query Match             32.8%;  Score 631; DB 10; Length 421;
Best Local Similarity   35.4%;  Pred. No. 4.8e-46;
Matches  131;  Conservative  68;  Mismatches  91;  Indels  80;  Gaps  2;

Qy    66 KKGKAQRQLYASAICLLFMIGEVVGGYLAHSLAVMTDAAHLLTDFASMLISLFSLWMSS 125
          :: ||||:|||:||: ||||:|.||||||| |:||||||||||:||| |||||||||
Db    50 ERSTSTKKLLIAVVLCIIFMSIEVFGVGGIEANSLAILTDAAHLLSDVAAYAISLFSLWASG 109

Qy   126 RPATKIMNFGWQRAEILGALVSVLSIWVVTGVLVYLAVERLISGDYEIDGGTMLITSGCA 185
          :||| ||| ||||||||||||:| :|||:| ||||| | |||| || |:|| ||| ||
Db   110 WEATPROSYGFFRIEILGALVSIQIIWLLAGILVYEAIERLINGTTEVHGFLMFIIAAFG 169

Qy   186 VAVNIIMGLTL------------------------HOSGHGHSH---------- 205
          :.|||   | |                         |||||||
Db   170 LLVNIAMALLGHDHSHGHGHDHGHGHSHGHDDAHEHSDHAHSHEDHGDLHTHGILHVL 229

Qy   206 ------------------------------GTTNQOEENPSVRAAFIHVI 225
                                         |::| :::|.| :.:||
Db   230 DHHHGEDSKGHADQLHGHETDQOTEPLLQOTCSEAEGDSKLGAKOKOQORNINMQGAYLHVL 289

Qy   226 GDFMOSMGVLIVAAYILYFKPEYKYVDPICFEVFSILVLGTTLTILRDVILVLMEGTPKGV 285
          ||.:.|:|| ||||:||:|||.|| || ||:|||| |||||  :||:|| ||||||| |
Db   290 GDSIOSVGVMICGAAIIWIKREWTIVDLICTLIFESVTVLGTTIRMLRNILEVLMESTPREI 349

Qy   286 DFTAVRDLLLSVEGVEALHSLHIWALTVAQPVLSVHIAIAQNTDAQAVLKTASSRLQGKF 345
          | |  : ||:|:| :| ||||||||:||.: :|| |: |  || |:|| :|: ||  |
Db   350 DATRLESGLCKMDEVIAVHELHIWAITVGKVLLACHVKIKRDANADMVLDKVVDYIRREY 409

Qy   346 HFHTVTIQIE 355
          :  ||||:|
Db   410 KISHVTIQVE 419
```

```
RESULT 18
ID   Q8T0G1        PRELIMINARY;     PRT;    498 AA.
AC   Q8T0G1;
DT   01-JUN-2002 (TrEMBLrel. 21, Created)
DT   01-JUN-2002 (TrEMBLrel. 21, Last sequence update)
DT   01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE   LD05335p.
GN   BEST:CK02137.
OS   Drosophila melanogaster (Fruit fly).
OC   Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC   Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC   Ephydroidea; Drosophilidae; Drosophila.
OX   NCBI_TaxID=7227;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=BERKELEY;
RA   Stapleton M., Brokstein P., Hong L., Agbayani A., Carlson J.,
RA   Champe M., Chavez C., Dorsett V., Farfan D., Frise E., George R.,
RA   Gonzalez M., Guarin H., Li P., Liao G., Miranda A., Mungall C.J.,
RA   Nunoo J., Pacleb J., Paragas V., Park S., Phouanenavong S., Wan K.,
RA   Yu C., Lewis S.E., Rubin G.M., Celniker S.;
RL   Submitted (DEC-2001) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AY069340; AAL39485.1; -.
SQ   SEQUENCE 498 AA; 55014 MW; AE0161C0DC3654E8 CRC64;

Query Match             31.5%;  Score 605; DB 5; Length 498;
Best Local Similarity   33.2%;  Pred. No. 9.2e-44;
Matches  129;  Conservative  76;  Mismatches  102;  Indels  81;  Gaps  5;

Qy    53 CHAQKGPDSHCDPKKG-KAQRQLYASAICLLFMIGEVVGGYLAHSLAVMTDAAHLLTDF 111
          |:||:|. |  :.:| :||:|||||:||| || |: ||||:| |||||||||||| :|
Db    53 CNHQPGFRANSKSKSAQEAKYKIMLAVALCCVFMIIEFLGGYVAGSLAIMTDAAHLLASDC 112

Qy   112 ASMLISLFSLWMSSRPATKTMNFGWQRAEILGALVSVLSIWVVTGVLVYLAVERLISGDY 171
          :|::|.| ::|: .|| |:.|||.::||:|||| :||:|||:| |:||: |:|| | .||
Db   113 ISFVIGLVAIWIGGRPPDERMSFGYKRFEVIGALASIIGIWFVTLLVVVAIERIFSQDF 172

Qy   172 EIDGGTMLITSGCAVAVNIIMGLTLHQS-----GHGHSH-------------------- 205
          ||||||||||||||||||||||:|||:      |||||
Db   172 ...
```

```
Qy   231  ELNADMMLISGIGIVINIVMMFVLHGSWFVNGNGHGHSHSHSHSHSHGNGHEPNDSL  290
              |::     ||     ||       :||    ||||||
Db   206  ------------------------------------GTTTN----QQE  212

Qy   291  SQTRSNSNFLTTIGSQSASTADEEDSIRKEINSNEHKIVITNGKKPTLTGTSNMELAHED  350
              :||||||||| ||||||||||||||||:||:|| |||| :|||:| |:   :|:|
Db   291  KNLNLRAAMIHVIGDLVQSIGVFLAAVLIKYCPGARXYADPLCTLIFSIIVIMTTLRLFRE  410

Qy   273  VILVLMEGTPKGVDFTAVRDLLSVEGVEALHSLHIWALTVAQPVLSVHIAIAQNTDAQA  332
              ::||||||||| |||||||||||||:||  |||||||| |||:  :|:| |:
Db   411  SLGIIVNAVPQNLNMRTLHLELGSIEGVRSLHHLNVWQQTSQQRVLMVHLVTDSRADGNE  470

Qy   333  VLKTASSRLQG-KFHFHTVTIQIEDYSE  359
              ||:||  ||| :||  || :||:|| :|
Db   471  VLQAATALVSSPRYNIKHSTIQLEPTTE  498
```

```
RESULT 19
ID   Q9M271            PRELIMINARY;        PRT;    334 AA.
AC   Q9M271;
DT   01-OCT-2000 (TrEMBLrel. 15, Created)
DT   01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT   01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE   Zinc transporter-like protein.
GN   F2I1F14.110.
OS   Arabidopsis thaliana (Mouse-ear cress).
OC   Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC   Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC   eurosids II; Brassicales; Brassicaceae; Arabidopsis.
OX   NCBI_TaxID=3702;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Choisne N., Robert C., Brottier P., Wincker P., Cattolico L.,
RA   Artiguenave F., Saurin W., Weissenbach J., Mewes H.W., Lemcke K.,
RA   Mayer K.F.X., Quetier F., Salanoubat M.;
RL   Submitted (FEB-2000) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AL138642; CAB71901.1; -
DR   InterPro; IPR002524; Cation_efflux.
DR   Pfam; PF01545; Cation_efflux; 1.
DR   TIGRFAMs; TIGR01297; CDF; 1.
SQ   SEQUENCE   334 AA;   37344 MW;  B0357A942A052E12 CRC64;
```

```
Qy    61  SHCDPKKGKA-QRQLVASAICLLFMIGEVVGGYLAHSLAVWTDAAHLLTDFASMLISLF  119
              |:    ||  ||||  |  ||||  ||  ||| || |||||:|||||| :|
Db    19  SSSDAKKRAASMRKLCFVVVLCLLFMSIEVVCGIKANSLAILADAAHLLTDVGAFAISML  78

Qy   120  SLWMSSRPATKTMNFGWQRAEILGALVSVLSIWVVTGVLVLYLAVERLI-SGDYEIDGGTM  178
              ||| ||:||  |:||   |:||||| |  :||:||||| :|  |||:  :||:  :|
Db    79  SLWASSWEANPRQSYGFFRIELGTLVSIQILIWLIIGILVVEAVTRLVQETNDDVDGFFM  138

Qy   179  LITSGCAVAVNIIMGLTL-HQSGHGHSHGTT------------NQEENPSVRAA  220
              | :|  |: |||| :||  |:| | |||| |              || :|:|  |
Db   139  VLVAARGLIVNIIMIVVLGHDHGHGHDHGHSHGYGERAEQLLEKSEIRNINVQGA  198

Qy   221  FIHVIGDFMQSMGVLVAAYILYFKPEYKYVDPICTFVFSILVLVLGTTLFILRDVILVLMEG  280
              ||| ||||||:||:|||| ||| ||  | :|| |  ||:|| :||||| :|  ||  |||
Db   199  VLVARGLIVNIIMIVVLGHDHGHGHDHGHSHGYGERAEQLLEKSEIRNINVQGA  258

Qy   281  TPKGVDFTAVRDLLLSVEGVEALHSLHIWALTVAQPVLSVHIAIAQNTDAQVLKTASSR  340
              || ||:| |:| |||||||||||||||||||||:|| ||||||||||||| ||:|||||
Db   259  TPREIDARQLEKGLMEIEEVDVHELHIWAITVGKALFSCHVKRPEAGDEMVLNKVIDY  318

Query Match
Best Local Similarity    31.1%;   Score 599;  DB 10;  Length 334;
Matches 126;  Conservative  64;  Mismatches 105;  Indels   20;  Gaps    4;
```

```
Qy   341  LQGKFHFHTVTIQIE  355
              ::|||||||||
Db   319  IWREYRISHVTIQIE  333
```

```
RESULT 20
ID   Q9V9I1            PRELIMINARY;        PRT;    452 AA.
AC   Q9V9I1;
DT   01-MAY-2000 (TrEMBLrel. 13, Created)
DT   01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT   01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE   CG11163 protein.
GN   BEST:CK02137 OR CG11163.
OS   Drosophila melanogaster (Fruit fly).
OC   Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC   Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC   Ephydroidea; Drosophilidae; Drosophila.
OX   NCBI_TaxID=7227;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=BERKELEY;
RX   MEDLINE=20196006; PubMed=10731132;
RA   Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA   Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
RA   George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA   Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
RA   Brandon R.C., Rogers Y.-H.C., Blazej R.G., Champe M., Pfeiffer B.D.,
RA   Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
RA   Abril J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
RA   Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA   Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
RA   Borkova D., Botchan M.R., Bouck J., Brokstein P., Brottier P.,
RA   Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA   Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
RA   de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA   Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA   Durbin K.J., Evangelista C.C., Ferraz C., Ferriera S., Fleischmann W.,
RA   Fosler C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,
RA   Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA   Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
RA   Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
RA   Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA   Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA   Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA   Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
RA   Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
RA   Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA   Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacleb J.M.,
RA   Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA   Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
RA   Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
RA   Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
RA   Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA   Wang Z.-Y., Wassarman D.A., Weinstock G.M., Weissenbach J.,
RA   Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
RA   Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA   Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
RA   Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
RT   "The genome sequence of Drosophila melanogaster.";
RL   Science 287:2185-2195(2000).
DR   EMBL; AE003785; AAF57308.1; -
DR   FlyBase; FBgn0025693; BEST:CK02137.
DR   InterPro; IPR002524; Cation_efflux.
DR   Pfam; PF01545; Cation_efflux; 1.
DR   TIGRFAMs; TIGR01297; CDF; 1.
SQ   SEQUENCE   452 AA;   49813 MW;  586380660608A88857 CRC64;

Query Match
Best Local Similarity    29.4%;   Score 565;  DB 5;  Length 452;
Matches 118;  Conservative  66;  Mismatches  88;  Indels   80;  Gaps    4;
```

```
RESULT 21
ID      Q92W23
AC      Q92W23;
DT      01-MAY-1999 (TrEMBLrel. 10, Created)
DT      01-MAY-1999 (TrEMBLrel. 10, Last sequence update)
DT      01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE      Putative zinc transporter.
GN      At2g29410.
OS      Arabidopsis thaliana (Mouse-ear cress).
OC      Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC      Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC      eurosids II; Brassicales; Brassicaceae; Arabidopsis.
OX      NCBI_TaxID=3702;
RN      [1]
RP      SEQUENCE FROM N.A.
RC      STRAIN=cv. COLUMBIA;
RX      MEDLINE=20083487; PubMed=10617197;
RA      Lin X., Kaul S., Rounsley S.D., Shea T.P., Benito M.-I., Town C.D.,
RA      Fujii C.Y., Mason T.M., Bowman C.L., Barnstead M.E., Feldblyum T.V.,
RA      Buell C.R., Ketchum K.A., Lee J.J., Ronning C.M., Koo H., Moffat K.S.,
RA      Cronin L.A., Shen M., VanAken S.E., Umayam L., Tallon L.J., Gill J.E.,
RA      Adams M.D., Carrera A.J., Creasy T.H., Goodman H.M., Somerville C.R.,
RA      Copenhaver G.P., Preuss D., Nierman W.C., White O., Eisen J.A.,
RA      Salzberg S.L., Fraser C.M., Venter J.C.;
RT      "Sequence and analysis of chromosome 2 of the plant Arabidopsis
RT      thaliana.";
RL      Nature 402:761-768(1999).
SQ      SEQUENCE   385 AA;  49A7B564F57FDC2B CRC64;
```

Query Match   26.0%; Score 500.5; DB 10; Length 385;
Best Local Similarity   29.8%; Pred. No. 7.2e-35;
Matches 114; Conservative 81; Mismatches 130; Indels 57; Gaps 5;

```
Qy      88  EVVGGYLAHSLAVMTDAAHLLTDFASMLISLFSLWMSSRPATKTMNFGWQRAEILGALVS 147
            : /|||: :|| : ||||:|||||:|||  :|| : :||||| |:| :|||:||:|||
Db     101  EFLGGYVAGSLAIMTDAAHLLASDCISFVIGLVAIWIGGRPPDERMSFGYKRFEVIGALAS 160

Qy     148  VLSIWVTGVLVYLAVERLISGDYEIDGGTGMLITSGCAVAVNIIMGLTLHQS------GH 201
            : /|:||: || : |:|:||||||||||||||:||||| ||: |:|:|| |        ||
Db     161  ILGIWFVTTLLVVVAIERIFSQDFELNADMMMLISGIVINIVMVMFVLHGSWFVNGNGH 220

Qy     202  GHSH----------------------------------------------------- 205
            |||
Db     221  GHSHSHSHSHSHSHGNGHEPNDSLSQTRSNSNFLTTIGSQSASTADEEDSIRKEINSNEH 280

Qy     206  ------------GTTN----QQEENPSVRAAFIHVIGDEFMQSMGVLVAAYILYFKPEYK 248
                        |:|:|    :||: || :|:||:|||:||||:||||||| ||:|||||
Db     281  KIVITNGKKPTLTGTSNMELAHEDKNLNLRAAMIHVIGDLVQSIGVFLAAVILIKVCPGAK 340

Qy     249  YVDPICFFVFESILVLGTTLTILRDVILVLMEGTPKGVDFTAVRDLLLSVEGVEALHSLHI 308
            |:||| :|: :|:| :||||| |||||:|||||||||||:|||||||||||||||| |||
Db     341  YADPLCTLLFISIIVIMTTLRLFRESLGIIVNAVPQNLNMRTLHLELGSSIEGVRSLHHLNV 400

Qy     309  WALTVAQPVLSVHIAIAQNTDAQAVLKTASSRLQG-KFHFHTVTIQIEDYSE 359
            : :|: :|||: :: :| : || : :||||||:|| : |||| |:|:|
Db     401  WQQTSQQRVLMVHLVTDSRADGNEVLQAATALVSSPRYNIKHSTIQLEPTTE 452
```

```
RESULT 22
ID      Q8SQT3
AC      Q8SQT3;
DT      01-JUN-2002 (TrEMBLrel. 21, Created)
DT      01-JUN-2002 (TrEMBLrel. 21, Last sequence update)
DT      01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE      Zinc transporter.
GN      ECU11_1510.
OS      Encephalitozoon cuniculi.
OC      Eukaryota; Microsporidia; Unikaryonidae; Encephalitozoon.
OX      NCBI_TaxID=6035;
RN      [1]
RP      SEQUENCE FROM N.A.
RC      STRAIN=GB-M1;
RP      SEQUENCE FROM N.A.
RC      Genoscope;
RL      Submitted (APR-2001) to the EMBL/GenBank/DDBJ databases.
RN      [2]
RP      SEQUENCE FROM N.A.
RC      STRAIN=GB-M1;
RX      MEDLINE=21576510; PubMed=11719806;
RA      Katinka M.D., Duprat S., Cornillot E., Metenier G., Thomarat F.,
RA      Prensier G., Barbe V., Peyretaillade E., Brottier P., Wincker P.,
RA      Delbac F., El Alaoui H., Peyret P., Saurin W., Gouy M.,
RA      Weissenbach J., Vivares C.P.;
RT      "Genome sequence and gene compaction of the eukaryote parasite
RT      Encephalitozoon cuniculi.";
RL      Nature 414:450-453(2001).
DR      EMBL; AL590450; CAD26061.1; -.
SQ      SEQUENCE   334 AA;  73FEAD2EE3FE8982 CRC64;
```

Query Match   22.5%; Score 432; DB 5; Length 334;
Best Local Similarity   32.9%; Pred. No. 4.7e-29;
Matches 102; Conservative 65; Mismatches 125; Indels 18; Gaps 7;

```
Qy      54  HAQKGPD--SHCDPKKGKAQRQLYVASAICLLFMIGEVVGGYLAHSLAVMTDAAHLLTDF 111
            |: :| | |   :|: :: |:|||||: :||||||||: ||||||||||||||||||||
Db       2  HGRKHPAQCSHVTCSNDADIRKISKVLLILMFPLELMGHWHWKTNSLSLLADSLHLLVDI 61

Qy     112  ASMLISLFSLWMSSRPATKTMNFGWQRAEILGALVSVLSIWVTGVLVYLAVERLISGDY 171
            :|||:|||||||||||| : |: |:||||: |: |||::|||| |||::|:||:||||:
Db      62  FGFIVSLLSLSWAKPSNRKMTFGYHRIEIIGSMWYSIGLIWAAVGYLIVIESFHKYLH-PA 120

Qy     172  EIDGGTMLITSGCAVA---VNIIMGLTLHQSGHGHGHSHGTTNQQEENPSVRAAFIHVIGDF 228
            |||| ||||||||||    |||:||||||| |||||:||  |  ||||||:|||||||||||
Db     121  EIDGGMFF---GIAVGFFVNCICIYLHYDEYQH-----KLKHKNLNIRATYVHVVGDL 172

Qy     229  MQSMGVLVAAYILYFKPEYKVDPICTFVFSILVLGTTLTILRDVILVLMEGTPKGVDFT 288
            : :|||::|||||||||| ||||| |||||: :| |||| |||||:|||||||||||||
Db     229  ... (229)
```

```
Db   173  IQSVGVIIAGMVTFYPSKAIVDVICTMFPSVLLISTGFVFRDGYIILAEGAPTDLDID 232
               :|:|||       |: |    |  :| |  ||  |: |   | |  ||| |  | :
Qy   289  AVRDLLLSVEGVEALHSLHIWALTVAQPVLSVHIAIAQN---TDAQVLKTASSRLQGKF 345
               |: :|| : :| | | ::|:|  |:|   |: |   |: |  ||| |  | :
Db   233  GMRSDIQDVENVYKIVDLYAWSISMNRSAVSIRI-LADDLLISDYENILLEVNHIIKGKY 291

Qy   346  HFHTVTIQIE 355
               |:|||
Db   292  LVDIVVVQID 301
```

RESULT 23

ID   Q9PF12   PRELIMINARY;   PRT;   321 AA.
AC   Q9PF12;
DT   01-OCT-2000 (TrEMBLrel. 15, Created)
DT   01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT   01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE   Cobalt-zinc-cadmium resistance protein.
GN   XF0866.
OS   Xylella fastidiosa.
OC   Bacteria; Proteobacteria; gamma subdivision; Xanthomonas group;
OC   Xylella.
OX   NCBI_TaxID=2371;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=20365717; PubMed=10910347;
RA   Simpson A.J.G., Reinach F.C., Arruda P., Abreu F.A., Acencio M.,
RA   Alvarenga M.H., Alves L.M.C., Araya J.E., Baia G.S., Baptista C.S.,
RA   Barros M.H., Bonaccorsi E.D., Bordin S., Bove J.M., Briones M.R.S.,
RA   Bueno M.R.P., Camargo A.A., Camargo L.E.A., Carraro D.M., Carrer H.,
RA   Colauto N.B., Colombo C.F., Costa F.F., Costa M.C.R., Costa-Neto C.M.,
RA   Coutinho L.L., Cristofani M., Dias-Neto E., Docena C., El-Dorry H.,
RA   Facincani A.P., Ferreira A.J.S., Ferreira V.C.A., Ferro J.A.,
RA   Fraga J.S., Franca S.C., Franco M.C., Frohme M., Furlan L.R.,
RA   Garnier M., Goldman G.H., Goldman M.H.S., Gomes S.L., Gruber A.,
RA   Ho P.L., Hoheisel J.D., Junqueira M.L., Kemper E.L., Kitajima J.P.,
RA   Krieger J.E., Kuramae E.E., Laigret F., Lambais M.R., Leite L.C.C.,
RA   Lemos E.G.M., Lemos M.V.F., Lopes S.A., Lopes C.R., Machado J.A.,
RA   Machado M.A., Madeira A.M.B.N., Madeira H.M.F., Marino C.L.,
RA   Marques M.V., Martins E.A.L., Martins E.M.F., Matsukuma A.Y.,
RA   Menck C.F.M., Miracca E.C., Miyaki C.Y., Monteiro-Vitorello C.B.,
RA   Moon D.H., Nagai M.A., Nascimento A.L.T.O., Netto L.E.S.,
RA   Nhani A. Jr., Nobrega F.G., Nunes L.R., Oliveira M.A.,
RA   de Oliveira M.C., de Oliveira R.C., Palmieri D.A., Paris A.,
RA   Peixoto B.R., Pereira G.A.G., Pereira H.A. Jr., Pesquero J.B.,
RA   Quaggio R.B., Roberto P.G., Rodrigues V., de Rosa A.J.M.,
RA   de Rosa V.E. Jr., de Sa R.G., Santelli R.V., Sawasaki H.E.,
RA   da Silva A.C.R., da Silva A.M., da Silva F.R., Silva W.A. Jr.,
RA   da Silveira J.F., Silvestri M.L.Z., Siqueira W.J., de Souza A.A.,
RA   de Souza A.P., Terenzi M.F., Truffi D., Tsai S.M., Tsuhako M.H.,
RA   Vallada H., Van Sluys M.A., Verjovski-Almeida S., Vettore A.L.,
RA   Zago M.A., Zatz M., Meidanis J., Setubal J.C.;
RT   "The genome sequence of the plant pathogen Xylella fastidiosa.";
RL   Nature 406:151-159(2000).
DR   EMBL; AE003926; AAF83676.1; -.
DR   InterPro; IPR002524; Cation_efflux.
DR   Pfam; PF01545; Cation_efflux; 1.
DR   TIGRFAMs; TIGR01297; CDF; 1.
KW   Complete proteome.
SQ   SEQUENCE   321 AA; 35412 MW; B57B7D1C925034DC CRC64;

Query Match   21.9%; Score 421.5; DB 16; Length 321;
Best Local Similarity   32.3%;
                        Matches 98; Conservative 64; Mismatches 126; Indels 15; Gaps   5;

```
Qy    53  CHAQKGPDSHCDPKKCKAQRQLYVASAICLLFMIGEVVGGYLAHSLAVMTDAAHLLTDFA 112
              :| | ||   :|  |||       | :  | :   | |     | |    :| |
Db    20  CH--MGHDHTHVPKGMHHERPLWWALGLTASFLVAEIIGAVLTRSLALLSDAVHMATDVF 77

Qy   113  SMLISLFSLWMSSRPATKTMNEFGWQRAEILGALVSVLSIWVTGVLVYLAVERLISGDYE 172
```

```
Db    78  ALMALVAVRLSRRPPDVRRTGYARLEAFGALVNGVLLFGVGGYILWEAVRRL-RAPQE 136
               :||| :    |: |      | :  | |     | |    :| |   || |
Qy   173  IDGGTMLITSGCAVAVNIIMGLTLHQSGHGHSHGTTNQQEENPSVRAAFIHVIGDFMQSM 232

Db   137  IVSSGMLVIAVLGLVINLIVMRLLH-AGRG-------------ENLSMKGAYLEVWSDMLGSV 185
               :| | ||
Qy   233  GVLVAAILYFKPEYKYVDPICTFVFSILVLGTTLTIIRDVILVLMEGTPKGVDFTAVRD 292

Db   186  AVIVAAIVIYVIGWY-WVDPVLAVLIGLWVLPRTWLLGEAVNVLLEGSVPKGFELLPIRD 244
               :|| :|| :||: | |  | | ::|:|   |: | : |   ||| |  | :
Qy   293  LLLSVEGVEALHSLHIWALTVAQPVLSVHIAIAQNTDAQAVLKTASSRLQGKFHFHVTI 352

Db   245  ALSHHPGVANVHDVHLWALGSRMPLLTAHVVALQGTAPDRLRRMLQELLHERFSIEHVTL 304
               |:|||
Qy   353  QIE 355
               |:|||
Db   305  QIE 307
```

RESULT 24

ID   Q9X7Q7   PRELIMINARY;   PRT;   312 AA.
AC   Q9X7Q7;
DT   01-NOV-1999 (TrEMBLrel. 12, Created)
DT   01-NOV-1999 (TrEMBLrel. 12, Last sequence update)
DT   01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE   Putative efflux protein.
GN   SCO6751 OR SC5F2A.34C.
OS   Streptomyces coelicolor.
OC   Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC   Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomyces.
OX   NCBI_TaxID=1902;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Oliver K., Harris D.;
RL   Submitted (APR-1999) to the EMBL/GenBank/DDBJ databases.
RN   [2]
RP   SEQUENCE FROM N.A.
RA   Bentley S.D., Parkhill J., Barrell B.G., Rajandream M.A.;
RL   Submitted (APR-1999) to the EMBL/GenBank/DDBJ databases.
RN   [3]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=97000351; PubMed=8843436;
RA   Redenbach M., Kieser H.M., Denapaite D., Eichner A., Cullum J.,
RA   Kinashi H., Hopwood D.A.;
RT   "A set of ordered cosmids and a detailed genetic and physical map for
RT   the 8 Mb Streptomyces coelicolor A3(2) chromosome.";
RL   Mol. Microbiol. 21:77-96(1996).
RN   [4]
RP   SEQUENCE FROM N.A.
RC   STRAIN=A3(2) / M145;
RA   Bentley S.D., Chater K.F., Cerdeno-Tarraga A.-M., Challis G.L.,
RA   Thomson N.R., James K.D., Harris D.E., Quail M.A., Kieser H.,
RA   Harper D., Bateman A., Brown S., Chandra G., Chen C.W., Collins M.,
RA   Cronin A., Fraser A., Goble A., Hidalgo J., Hornsby T., Howarth S.,
RA   Huang C.-H., Kieser T., Larke L., Murphy L., Oliver K., O'Neil S.,
RA   Rabbinowitsch E., Rajandream M.A., Rutherford K., Rutter S.,
RA   Seeger K., Saunders D., Sharp S., Squares R., Squares S., Taylor K.,
RA   Warren T., Wietzorrek A., Woodward J., Barrell B.G., Parkhill J.,
RA   Hopwood D.A.;
RT   "Complete genome sequence of the model actinomycete Streptomyces
RT   coelicolor A3(2).";
RL   Nature 417:141-147(2002).
DR   EMBL; AL049587; CAB40701.1; -.
DR   InterPro; IPR002524; Cation_efflux.
DR   Pfam; PF01545; Cation_efflux; 1.
DR   TIGRFAMs; TIGR01297; CDF; 1.
SQ   SEQUENCE   312 AA; 32579 MW; 1F0F44DFB326ADA5 CRC64;

Query Match   21.7%; Score 416.5; DB 16; Length 312;

```
Best Local Similarity 32.5%; Pred. No. 9.3e-28;
Matches 106; Conservative  66; Mismatches 125; Indels 29; Gaps 8;

Qy   52 HCHAQKGPDSHCDPKKGKQAQR-QLYVASAICLLFMIGEVVGGYIAHSLAVMTDAAHLLTD 110
            |:| :|:|||   :||  ::|||  :: ||  ||   :|||  |: ::|||| | |
Db    7 HGHAHGG---HAGGTATAAYRGLRVALSITLLTVMVVELVGGLVADSLALIADAAHMATD 63

Qy  111 FASMLISLFSLWMSSRPATKTMNFGWQRAEILGALVSVLSIWVVTGVLVYLAVERLISGD 170
           :: :||| ||   |||| :  | ||||: |  :  || :|  :|  || | : :
Db   64 ALGLGMALLAVHFASRPSDRRFFGYARAEILAALANCLLLGVGGYVLYEAIDRFVTPA 123

Qy  171 YEIDGGTMLITSGCAVAVNIIMGLTLHQSGHGSHGTTNQQEENPSVRAAFIHVIGDEMQ 230
           |  |  |:| | | :|: |  :|| | |   |  ||||  | ||  :|  |:| |  |
Db  124 DIYAAGPTVVF--GAIGLVANMVSLSLLMRG------QKESLNVRGAFLEVVADALG 171

Qy  231 SMGVLVAAYILYFKPEYKYVDPICTEVFSILVLGTTVLTILRDVILVLMEGTPKGVDFTAV 290
          |  : | :| |     : ||| | |::|: | |||| :|||||::|:||||||||| ||
Db  172 SLAVIVSALVI-LATGWQPADPIASLLIGLMIVPRTLRLLREILDVLLEAAPKGVDIAEV 230

Qy  291 RDLLLSVEGVEALHSLHIWALTVAQPVLSVHIAIAQNTDAQAVLKTASSR----LQG--- 343
          | ::| | ||:  ||| |: | :  ||||:|   : | :|  | |:   :    | |
Db  231 RAHILALDGVEDVHDLHAWTITSGMPVLSAHVVV----DGEALSAIGHEKMLHELQGCLG 286

Qy  344 -KFHFHTVTIQIEDYSEDMKDCQACQ 368
           :| | ::  :| |  :   | || |
Db  287 DHFDVEHCTFQLEPSGHAEHEARLCR 312
```

RESULT 25
Q8Y480

ID   Q8Y480;   PRELIMINARY;   PRT;   303 AA.
AC   Q8Y480;
DT   01-MAR-2002 (TrEMBLrel. 20, Created)
DT   01-MAR-2002 (TrEMBLrel. 20, Last sequence update)
DT   01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE   Hypothetical protein lmo2575.
GN   LMO2575.
OS   Listeria monocytogenes.
OC   Bacteria; Firmicutes; Bacillus/Clostridium group; Bacillales;
OC   Listeriaceae; Listeria.
OX   NCBI_TaxID=1639;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=EGD-E / SEROVAR 1/2A;
RX   MEDLINE=21537279; PubMed=11679669;
RA   Glaser P., Frangeul L., Buchrieser C., Rusniok C., Amend A.,
RA   Baquero F., Berche P., Bloecker H., Brandt P., Chakraborty T.,
RA   Charbit A., Chetouani F., Couve E., de Daruvar A., Dehoux P.,
RA   Domann E., Dominguez-Bernal G., Duchaud E., Durant L., Dussurget O.,
RA   Entian K.-D., Fsihi H., Garcia-del Portillo F., Garrido P.,
RA   Gautier L., Goebel W., Gomez-Lopez N., Hain T., Hauf J., Jackson D.,
RA   Jones L.-M., Kaerst U., Kreft J., Kuhn M., Kunst F., Kurapkat G.,
RA   Madueno G., Maitournam A., Mata Vicente J., Ng E., Nedjari H.,
RA   Nordsiek G., Novella S., de Pablos B., Perez-Diaz J.-C., Purcell R.,
RA   Remmel B., Rose M., Schlueter T., Simoes N., Tierrez A.,
RA   Vazquez-Boland J.-A., Voss H., Wehland J., Cossart P.;
RT   "Comparative genomics of Listeria species.";
RL   Science 294:849-852(2001).
DR   EMBL; AL591983; CAD00653.1; -.
DR   ListList; LMO2575; -.
DR   InterPro; IPR002524; Cation_efflux.
DR   Pfam; PF01545; Cation_efflux; 1.
DR   TIGRFAMs; TIGR01297; CDF; 1.
KW   Hypothetical protein; Complete proteome.
SQ   SEQUENCE 303 AA; 33140 MW; BE32A94EA3FB81FE CRC64;

```
Query Match  21.6%; Score 414.5; DB 16; Length 303;
Best Local Similarity 30.1%; Pred. No. 1.3e-27;
Matches 93; Conservative 72; Mismatches 127; Indels 17; Gaps 5;

Qy   47 AQSNHHCHAQKGPDSHCDPKKGKQAQRQLYVASAICLLFMIGEVVGGYIAHSLAVMTDAAH 106
          :: | | || || |: ::|     :|||||:: |||||||  :| ||:| |||| || ||
Db    2 AHNHDHAHGHNHNHAHNANKK-----SLFISFILIATFMVVEVIGGIMTNSLALLSDAGH 56

Qy  167 ISGDYEIDGGTMLITSGCAVAVNIIMGLTLHQSGHGSHGTTNQQEENPSVRAAFIHVIG 226
```

```
Qy  107 LLTDFASMLISLFSLWMSSRPATKTMNFGWQRAEILGALVSVLSIWVVTGVLVYLAVERL 166
          ||:| ||: :|||||||||||||||||||||||||||||:|||||||||||||||||:||
Db   57 MLSDAVALGLSLAAFKFGERAASSDKTYGYKRFEILAAFLNGLTLLVGISVPFIFYEAIGRF 116

Qy  167 ISGDYEIDGGTMLITSGCAVAVNIIMGLTLHQSGHGSHGTTNQQEENPSVRAAFIHVIG 226
          | || | ||||: | :| |  | | |: | ::    :    : :    :|  ||:||:|
Db  117 FDPPQVIGAGMWTI-SVIGLLINILVAWILMK------GDTS--ENLNRSAFLHVLG 165

Qy  227 DEMQSMGVLVAAYILYFKPEYKYVDPICTEVFSILVLGTTVLTILRDVILVLMEGTPKGVD 286
          | : | |::  |: : | |  :::||::|: : :| :| :| | | :|: ||||| | :|
Db  166 DLLGSVGAITAALLIFL-GWNIADPIASVLVVAALLILVSGMRVLKDAIHILMEGKPANVD 224

Qy  287 FTAVRDLLLSVEGVEALHSLHIWALTVAQPVLSVHIAIAQNTDAQAVLKTASSRLQGKFH 346
           | :: ||  |  |::|| ||||| || ||  || :  |: |   : |    ||||: :|
Db  225 TEEIKTFFQQQDGVKEVHDLHWAITSDFNALSAHLTVCEDADRDKILADIEHYLQENFS 284

Qy  347 FHTVTIQIE 355
          :| :||:|
Db  285 LEHSTIQLE 293
```

RESULT 26
Q927R8

ID   Q927R8;   PRELIMINARY;   PRT;   303 AA.
AC   Q927R8;
DT   01-DEC-2001 (TrEMBLrel. 19, Created)
DT   01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT   01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE   Hypothetical protein lin2720.
GN   LIN2720.
OS   Listeria innocua.
OC   Bacteria; Firmicutes; Bacillus/Clostridium group; Bacillales;
OC   Listeriaceae; Listeria.
OX   NCBI_TaxID=1642;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=CLIP 11262 / SEROVAR 6A;
RX   PubMed=11679669;
RA   Glaser P., Frangeul L., Buchrieser C., Rusniok C., Amend A.,
RA   Baquero F., Berche P., Bloecker H., Brandt P., Chakraborty T.,
RA   Charbit A., Chetouani F., Couve E., de Daruvar A., Dehoux P.,
RA   Domann E., Dominguez-Bernal G., Duchaud E., Durant L., Dussurget O.,
RA   Entian K.-D., Fsihi H., Garcia-del Portillo F., Garrido P.,
RA   Gautier L., Goebel W., Gomez-Lopez N., Hain T., Hauf J., Jackson D.,
RA   Jones L.-M., Kaerst U., Kreft J., Kuhn M., Kunst F., Kurapkat G.,
RA   Madueno G., Maitournam A., Mata Vicente J., Ng E., Nedjari H.,
RA   Nordsiek G., Novella S., de Pablos B., Perez-Diaz J.-C., Purcell R.,
RA   Remmel B., Rose M., Schlueter T., Simoes N., Tierrez A.,
RA   Vazquez-Boland J.-A., Voss H., Wehland J., Cossart P.;
RT   "Comparative genomics of Listeria species.";
RL   Science 294:849-852(2001).
DR   EMBL; AL596173; CAC97946.1; -.
DR   ListList; LIN2720; -.
DR   InterPro; IPR002524; Cation_efflux.
DR   Pfam; PF01545; Cation_efflux; 1.
DR   TIGRFAMs; TIGR01297; CDF; 1.
KW   Hypothetical protein; Complete proteome.
SQ   SEQUENCE 303 AA; 33165 MW; AAC6D63CC5473422 CRC64;

```
Query Match  21.3%; Score 410.5; DB 16; Length 303;
Best Local Similarity 30.4%; Pred. No. 2.9e-27;
Matches 94; Conservative 70; Mismatches 128; Indels 17; Gaps 5;

Qy   47 AQSNHHCHAQKGPDSHCDPKKGKQAQRQLYVASAICLLFMIGEVVGGYIAHSLAVMTDAAH 106
          :: | | || || |: ::|     :|||||:: |||||||  :| ||:| |||| || ||
Db    2 AHNHDHAHGHNHNHTHNANKK-----SLFISFILIATFMVVEVIGGIMTNSLALLSDAGH 56

Qy  107 LLTDFASMLISLFSLWMSSRPATKTMNFGWQRAEILGALVSVLSIWVVTGVLVYLAVERL 166
          ||:| ||: :|||||||||||||||||||||||||||||:|||||||||||||||||:||
Db   57 MLSDAVALGLSLAAFKFGERAASSDKTYGVKRFEILAAFLINGLTLVGISVFIFYEAIGRF 116

Qy  167 ISGDYEIDGGTMLITSGCAVAVNIIMGLTLHQSGHGSHGTTNQQEENPSVRAAFIHVIG 226
```

•

```
Db   117 FDPPQVIGAGMMTI-SVIGLLINILVAWILMK------GDTS---ENLNMRSAFLHVLG 165
             : ::|| :  |  :|::        :    |||          |  :  |:|:|::||
Qy   227 DFMQSMGVLVAAYILYFKPEYKYVDPICTFVFSILVLGTTLTLILRDVILVLMEGTPKGVD 286
             ::|:  ||  | || | ||| |:   ||||||||||| ::||:::: |||| | | |
Db   166 DLLGSVGAITAALLIFF-GWNIADPIASVIVAAILVSGWRVLKDAIHILMEGKPANVD 224

Qy   287 FTAVRDLLLSVEGVEALHSLHIWALTVAQPVLSVHIAIAQNTDAQAVLKTASSRLQGKFH 346
             : || |::::  | | :: :| |  :|:|| : :|  ||    ||| |  ||
Db   225 TEEIKTFFQEQDGVTEVHDLHVWAITSDFNALTFAHLTVAEDADRDKILITDIEHYLQENFS 284

Qy   347 FHTVTIQIE 355
             |||:||:|:|
Db   285 LEHSTIQLE 293
```

RESULT 27
ID   Q9R494   PRELIMINARY;   PRT;   316 AA.
AC   Q9R494;
DT   01-MAY-2000 (TrEMBLrel. 13, Created)
DT   01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT   01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE   ORF 2 upstream of GlGB coding region.
OS   Bacillus stearothermophilus.
OC   Bacteria; Firmicutes; Bacillus/Clostridium group; Bacillales;
OC   Geobacillus.
OX   NCBI_TaxID=1422;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=92079888; PubMed=1745226;
RA   Kiel J.A., Boels J.M., Beldman G., Venema G.;
RT   "Molecular cloning and nucleotide sequence of the glycogen branching
RT   enzyme gene (glgB) from Bacillus stearothermophilus and expression in
RT   Escherichia coli and Bacillus subtilis.";
RL   Mol. Gen. Genet. 230:136-144(1991).
RT   InterPro; IPR002524; Cation_efflux.
DR   Pfam; PF01545; Cation_efflux; 1.
DR   TIGRFAMS; TIGR01297; CDF; 1.
SQ   SEQUENCE 316 AA; 35833 MW; 2025311569707A23 CRC64;

Query Match   21.3%; Score 409.5; DB 2; Length 316;
Best Local Similarity   28.48%;
  Matches   82; Conservative   86; Mismatches 110; Indels 11; Gaps 3; Pred. No. 3.8e-27;

```
Qy   67  KGKAQRQLYVASAICLLFMIGEVVGGYILAHSLAVMTDAAHLLTDFASMLISLFSLWMSSR 126
             ::|:| |:| | ||| ||||||||||| |||:|||| |||:||  :|:|||||:|||:||
Db   16  QNQSKKLIWITILVLTLFFTIVEIVGGLLSNSLLALLSDSAHMASDVLALGLSMIALYLAMR 75
Qy   127 PATKTMNFPGWQRAEILGALVSVLSIWVVTGVLVLYLAVERLISGDYEIDGGTMLITSGCAV 186
             :  :  :|: ||| :|:: |: :| :|  ||:| :|| ||:| ::  |::
Db   76  PPNHRFTFGYLRFEIITSFLNGLTLALIISIGILWEGIQRFISPE-PIDPRLMLTISSIGF 134
Qy   187 AVNIIMGLTLHQSGHGHSHGTTINQQEENPSVRAAFIHVIGDFMQSMGVLVAAYILYFKPE 246
             :|| |:|::| |        ::|| ||:: |  |  :|:||:|   | |:::|| :|| |
Db   135 IVNLILTILSRS-------TKEEENLNIKSALWHFIGDLLSSIGVIISAILIYFTGF 185
Qy   247 YKVVDPICTFVFSILVLGTTLTLILRDVILVLMEGTPKGVDFTAVRDLLLSVEGVEALHSL 306
             | :|| :  :: || :|  |  :|  :| :|||| || |:| | :|  |:| | | ||:|
Db   186 Y-FFDPLISIIIAAIIFTGGAKIIRESYLILMEAVPDQFNLDQIRGDIRQIEGVEDVHDM 244
Qy   307 HIWALTVAQPVLSVHIAIAQNTDAQAVLKTASSRLQGKFHFHTVTIQIE 355
             |:||: : :  |:| || : | | :  |: :: :: :: || :||:|:|
Db   245 HLWAVSTDHYSLFAHVFISEHIQPFCVILAINEMLNKKYGIKHTTIQVE 293
```

RESULT 28
ID   Q99X53   PRELIMINARY;   PRT;   319 AA.
AC   Q99X53;
DT   01-JUN-2001 (TrEMBLrel. 17, Created)
DT   01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
DT   01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE   Hypothetical protein SAV0168.
GN   SAV0168 OR SA0163.
OS   Staphylococcus aureus (strain Mu50 / ATCC 700699), and
OS   Staphylococcus aureus (strain N315).
OC   Bacteria; Firmicutes; Bacillus/Clostridium group; Bacillales;
OC   Staphylococcus.
OX   NCBI_TaxID=158878, 158879;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   SPECIES=S. aureus (strain Mu50), and S. aureus (strain N315);
RX   MEDLINE=21311952; PubMed=11418146;
RA   Kuroda M., Ohta T., Uchiyama I., Baba T., Yuzawa H., Kobayashi I.,
RA   Cui L., Oguchi A., Aoki K.-I., Nagai Y., Lian J.-Q., Ito T.,
RA   Kanamori M., Matsumaru H., Maruyama A., Murakami H., Hosoyama A.,
RA   Mizutani-Ui Y., Takahashi N.K., Sawano T., Inoue R.-I., Kaito C.,
RA   Sekimizu K., Hirakawa H., Kuhara S., Goto S., Yabuzaki J.,
RA   Kanehisa M., Yamashita A., Oshima K., Furuya K., Yoshino C., Shiba T.,
RA   Hattori M., Ogasawara N., Hayashi H., Hiramatsu K.;
RT   "Whole genome sequencing of meticillin-resistant Staphylococcus
RT   aureus.";
RL   Lancet 357:1225-1240(2001).
DR   EMBL; AP003358; BAB56330.1; -.
DR   EMBL; AP003129; BAB41383.1; -.
DR   InterPro; IPR002524; Cation_efflux.
DR   Pfam; PF01545; Cation_efflux; 1.
DR   TIGRFAMS; TIGR01297; CDF; 1.
KW   Hypothetical protein; Complete proteome.
SQ   SEQUENCE 319 AA; 36235 MW; AAC27E49D5F3157 CRC64;

Query Match   20.8%; Score 399.5; DB 16; Length 319;
Best Local Similarity   30.1%;
  Matches   93; Conservative   84; Mismatches 109; Indels 23; Gaps 6; Pred. No. 2.8e-26;

```
Qy   51  HHCHAQKGPDSHCDPKKGKAQRQLYVASAICLLFMIGEVVGGYILAHSLAVMTDAAHLLTD 110
             || :| |:   :    ::|: |  :|  | |||||||||||| ||:||| | | |  |:
Db   16  HHVNHQKIQQS------SKKTLWASLIITLFTVIEFVGGLVSNSLLALLSDSFHMLSD 67
Qy   111 FASMLISLFSLWMSSRPATKTMNFPGWQRAEILGALVSVLSIWVVTGVLVLYLAVERLISGD 170
             :|:|:|| |        |: ||:: |   |::|  |:: :| |:  |  |:||  :|:
Db   68  VLALGLSMIALYFASKKPTARYTFGYLRFEILVAFLNGLALIVISIWILYEAIVRIIY-P 126
Qy   171 YEIDGGTMLITSGCAVANIIMGLTLHQSGHGHSHGTTINQQEENPSVRAAFIHVIGDFMQ 230
             :  |: ::  :| ::: |  :|  |          :|| ||:::   | :|::|:|:|
Db   127 QPIESGIMFMIASIGLLVNIILTILVRS------LKQEDNINIQSALWHFMGDLLN 177
Qy   231 SMGVLVAAYILYFKPEYKYVDPICTFVFSILVLGTTLTLILRDVILVLMEGTPKGVDFTAV 290
             | |::|| ::|| |   :||| :|   || ::|  |   |: :| :||| | | | | :
Db   178 SIGVIVAVLIYF-TGWRIIDPIISIVISLILRGGYKITRNAWLILMESVPQHLDTDQI 236
Qy   291 RDLLLSVEGVEALHSLHIWALTVAQPVLSVHIAI----AQNTDAQAVLKTASSRLQGKFHF 347
             :|:  | :  :  |  |:| :: : |:|:   |       :  ||   | || :: ::|:
Db   237 MADIKNIDGILDVHEFHLWSITTEHYSLSAHVVLDKKYEGDDYQAI-DQVSSLLKEKYGI 295
Qy   348 HTVTIQIED 356
             |::|:||:|
Db   296 AHSTLQIEN 304
```

RESULT 29
ID   007084   PRELIMINARY;   PRT;   311 AA.
AC   007084;
DT   01-JUL-1997 (TrEMBLrel. 04, Created)
DT   01-JUL-1997 (TrEMBLrel. 04, Last sequence update)
DT   01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE   Cation transport protein YRDO.
GN   YRDO OR CZCD.
OS   Bacillus subtilis.
OC   Bacteria; Firmicutes; Bacillus/Clostridium group; Bacillales;
OC   Bacillaceae; Bacillus.
OX   NCBI_TaxID=1423;
RN   [1]
RP   SEQUENCE FROM N.A.

```
RC   STRAIN=168;
RA   Sorokin A., Bolotin A., Purnelle B., Hilbert H., Lauber J.,
RA   Duesterhoeft A., Ehrlich S.D.;
RL   Submitted (MAY-1997) to the EMBL/GenBank/DDBJ databases.
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=168;
RX   MEDLINE=98044033; PubMed=9384377;
RA   Kunst F., Ogasawara N., Moszer I., Albertini A.M., Alloni G.,
RA   Azevedo V., Bertero M.G., Bessieres P., Bolotin A., Borchert S.,
RA   Borriss R., Boursier L., Brans A., Braun M., Brignell S.C., Bron S.,
RA   Brouillet S., Bruschi C.V., Caldwell B., Capuano V., Carter N.M.,
RA   Choi S.K., Codani J.J., Connerton I.F., Cummings N.J., Daniel R.A.,
RA   Denizot F., Devine K.M., Duesterhoeft A., Ehrlich S.D., Emmerson P.T.,
RA   Entian K.D., Errington J., Fabret C., Ferrari E., Foulger D.,
RA   Fritz C., Fujita M., Fujita Y., Fuma S., Galizzi A., Galleron N.,
RA   Ghim S.Y., Glaser P., Goffeau A., Golightly E.J., Grandi G.,
RA   Guiseppi G., Guy B.J., Haga K., Haiech J., Harwood C.R., Henaut A.,
RA   Hilbert H., Holsappel S., Hosono S., Hullo M.F., Itaya M., Jones L.,
RA   Joris B., Karamata D., Kasahara Y., Klaerr-Blanchard M., Klein C.,
RA   Kobayashi Y., Koetter P., Koningstein G., Krogh S., Kumano M.,
RA   Kurita K., Lapidus A., Lardinois S., Lauber J., Lazarevic V.,
RA   Lee S.M., Levine A., Liu H., Masuda S., Mauel C., Medigue C.,
RA   Medina N., Mellado R.P., Mizuno M., Moestl D., Nakai S., Noback M.,
RA   Noone D., O'Reilly M., Ogawa K., Ogiwara A., Oudega B., Park S.H.,
RA   Parro V., Pohl T.M., Portetelle D., Porwollik S., Prescott A.M.,
RA   Presecan E., Pujic P., Purnelle B., Rapoport G., Rey M., Reynolds S.,
RA   Rieger M., Rivolta C., Rocha E., Roche B., Rose M., Sadaie Y.,
RA   Sato T., Scanlan E., Schleich S., Schroeter R., Scoffone F.,
RA   Sekiguchi J., Sekowska A., Seror S.J., Serror P., Shin B.S., Soldo B.,
RA   Sorokin A., Tacconi E., Takagi T., Takahashi H., Takemaru K.,
RA   Takeuchi M., Tamakoshi A., Tanaka T., Terpstra P., Tognoni A.,
RA   Tosato V., Uchiyama S., Vandenbol M., Vannier F., Vassarotti A.,
RA   Viari A., Wambutt R., Wedler E., Wedler H., Weitzenegger T.,
RA   Winters P., Wipat A., Yamamoto H., Yamane K., Yasumoto K., Yata K.,
RA   Yoshida K., Yoshikawa H.F., Zumstein E., Yoshikawa H., Danchin A.;
RT   "The complete genome sequence of the gram-positive bacterium Bacillus
RT   subtilis.";
RL   Nature 390:249-256(1997).
RN   [3]
RP   SEQUENCE FROM N.A.
RC   STRAIN=168;
RA   Kunst F., Ogasawara N., Yoshikawa H., Danchin A.;
RL   Submitted (NOV-1997) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; U93876; AAB80907.1; -.
DR   EMBL; Z99117; CAB14606.1; -.
DR   InterPro; IPR002524; -.
DR   Pfam; PF01545; Cation_efflux; 1.
DR   TIGRFAMs; TIGR01297; CDF; 1.
KW   Complete proteome.
SQ   SEQUENCE 311 AA; 34162 MW; 9F495E9DF8E96DB0 CRC64;

Query Match             20.7%;  Score 397.5;  DB 16;  Length 311;
Best Local Similarity   28.2%;  Pred. No. 4e-26;
Matches   86;  Conservative   86;  Mismatches 118;  Indels  15;  Gaps  5;

Qy     67 KGKAQRQLYVASAICLLFMIGEVVGGYLAHSLAVMTDAAHLLTDFASMLISLFSLWMSSR 126
           : :  ::  :   ::   :  :  ::   ::: :: :::  :::::::::  :  :: :
Db      7 EGANKKVLLISFIMITGYMIIEAIGGFLTNSLALLSDAGHMLSDSISLMVALIAFTLAEK 66

Qy    127 PATKTMNFPGWQRAEILGALVSVLSIWVTGVLVYLAVERLIGSDYEIDGGTMLITSGCAV 186
           : :  ::   ::: :::::  ::: : :: :  :: :  : ::  ::  :::::::
Db     67 KANHNKTFPGYKRFEILAAVINGAALILSLYIIYEAIER-FSNPPKVATTGMLFISIGCL 125

Qy    187 AVNIIMGLTLHQSGHGHSHGTNQQEENPSVRAAFIHVIGDFMQSMGVLVAVILYFKPE 246
           :::  ::::::  :: ::  : ::: : ::::  : :  :  ::::  :::  :
Db    126 VVNLLVAWIMMSGG--------DTKNNLNIRGAYLHVISDMLGSVGAILAAILIFF-G 175

Qy    247 YKVVDPICTFVFSFLVLGGTTLTLIRDVLIVLVLMEGTPKGVDFTAVRDLLLSVEGVEALHSL 306
           :| :: :: :::::  :   : ::  ::  : :::  :   ::   ::::  :   :
Db    176 WGNADPLASIVAILVLRSGYNVTKDSTHILMEGTPENIDVSDIRTIEGTEGTQNIHDL 235

Qy    307 HIWALTVAQPVLSVHIAIAQN--TDAQAVLKTASSRLQGKFHFHTVIQIEDYSEDMKD 363
           ::|::: :  : :  ::: : :  :|:: :::: ::::: | : | :
Db    236 HIWSITSGLNALSCHAVVDDQQLTISESENIIRKIEHELEHKGITH-VTIQMETEAHNHDN 294

Qy    364 CQACQ 368
           :: ::
Db    295 AILCQ 299


RESULT 30
ID   Q8YFT2      PRELIMINARY;      PRT;   301 AA.
AC   Q8YFT2;
DT   01-MAR-2002 (TrEMBLrel. 20, Created)
DT   01-MAR-2002 (TrEMBLrel. 20, Last sequence update)
DT   01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE   Cobalt-zinc-cadmium resistance protein CzcD.
GN   BMEI1438.
OS   Brucella melitensis.
OC   Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group;
OC   Brucellaceae; Brucella.
OX   NCBI_TaxID=29459;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=16M / ATCC 23456 / BIOTYPE 1;
RX   MEDLINE=20020109; PubMed=11756688;
RA   DelVecchio V.G., Kapatral V., Redkar R.J., Patra G., Mujer C., Los T.,
RA   Ivanova N., Anderson I., Bhattacharyya A., Lykidis A., Reznik G.,
RA   Jablonski L., Larsen N., D'Souza M., Bernal A., Mazur M., Goltsman E.,
RA   Selkov E., Elzer P.H., Hagius S., O'Callaghan D., Letesson J.-J.,
RA   Haselkorn R., Kyrpides N., Overbeek R.;
RT   "The genome sequence of the facultative intracellular pathogen
RT   Brucella melitensis.";
RL   Proc. Natl. Acad. Sci. U.S.A. 99:443-448(2002).
DR   EMBL; AE009580; AAL52619.1; -.
DR   InterPro; IPR002524; -.
DR   Pfam; PF01545; Cation_efflux; 1.
DR   TIGRFAMs; TIGR01297; CDF; 1.
KW   Complete proteome.
SQ   SEQUENCE 301 AA; 32663 MW; 268B9E962EB7CE32 CRC64;

Query Match             20.5%;  Score 393.5;  DB 16;  Length 301;
Best Local Similarity   30.9%;  Pred. No. 8.5e-26;
Matches   93;  Conservative   66;  Mismatches 123;  Indels  19;  Gaps  6;

Qy     58 GPD-SHCDPKKGKAQRQLYVASAICLLFMIGEVVGGYLAHSLAVMTDAAHLLTDFASMLI 116
           :   ::  :    |::    | :|| ||| :  ::  || : : |||||| |||  :||
Db      6 GADHEHADVKNTPISR-LWIAFGLGIFMIAEVIGSFATGSLALLSDAMHHATDAFALL 64

Qy    117 SLFSLWMSSRPATKTMNFPGWQRAEILGALVSVLSIWVTGVLVYLAVERLIGSDYEIDG 176
           :| :   ::||| :    | ||| ||||::  |: : | : |  ||:||:  |  : :
Db     65 ALIAIYLGRRPADVLRTYGYARFEILAAANALLLGVAFYILYEAWERL-SEPADVQSL 123

Qy    177 TMLITSGCAVAVNIIMG--LTLHQSGHGHSHGTNQQEENPSVRAAFIHVIGDFMQSMGV 234
           ||| :| |  | ||:    |||  ::      : :::  | | |||: :| |:  |:||
Db    124 GMLAVAIVGLVMNFISMRILTVH-------------KDDSLNVKGAYLEVWADMLGSVGV 170

Qy    235 LVAAYILYFKPEYKVVDPICTFVFSILVLGGTTLTLIRDVLIVLVLMEGTPKGVDFTAVRDLL 294
           : ||:|:||   |  ||  | | | |:| ::    :  :: ||  : :: | | |  :
Db    171 IAAAIIYL-TGWEWVDSAIAVGIGFMVFPRTWVLLKECINILLEGVPAGGDVKKLEAAI 229

Qy    295 LSVEGVEALHSLHIWALTVAQPVLSVHIAIAQNTDAQAVLKTASSRLQGKFHFHTVIQI 354
           | | ||  : ||| |:|| :  ::   ::: ||:| :  ::  :| | : |  | :
Db    230 LAVPGVGASVHDLHVWSLTKTEHSLTAHLVLAQEEADCETVRRAVEHVLQNDYDLHITTLQT 289

Qy    355 E 355
           :
Db    290 E 290


RESULT 31
ID   Q05214      PRELIMINARY;      PRT;   317 AA.
AC   Q05214;
```

DT   01-JUL-1997 (TrEMBLrel. 04, Created)
DT   01-JUL-1997 (TrEMBLrel. 04, Last sequence update)
DT   01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE   Membrane protein.
OS   Bacillus stearothermophilus.
OC   Bacteria; Firmicutes; Bacillus/Clostridium group; Bacillales;
OC   Geobacillus.
OX   NCBI_TaxID=1422;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   STRAIN=TRBE14;
RA   Takata H.;
RL   Submitted (AUG-1996) to the EMBL/GenBank/DDBJ databases.
RN   [2]
RP   SEQUENCE OF 1-64 FROM N.A.
RX   STRAIN=TRBE14;
RX   MEDLINE=95031021; PubMed=7944355;
RA   Takata H., Takaha T., Kuriki T., Okada S., Takagi M., Imanaka T.;
RT   "Properties and active center of the thermostable branching enzyme
RT   from Bacillus stearothermophilus.";
RL   Appl. Environ. Microbiol. 60:3096-3104(1994).
RN   [3]
RP   SEQUENCE FROM N.A.
RX   STRAIN=TRBE14;
RX   MEDLINE=97386405; PubMed=9244254;
RA   Takata H., Takaha T., Okada S., Takagi M., Imanaka T.;
RT   "Characterization of a gene cluster for glycogen biosynthesis and a
RT   heterotetrameric ADP-glucose pyrophosphorylase from Bacillus
RT   stearothermophilus.";
RL   J. Bacteriol. 179:4689-4698(1997).
DR   EMBL; D87026; BAA19587.1; -.
DR   InterPro; IPR002524; Cation_efflux.
DR   Pfam; PF01545; Cation_efflux; 1.
DR   TIGRFAMs; TIGR01297; CDF; 1.
SQ   SEQUENCE   317 AA;   35468 MW;   D3739EE34C7310F2 CRC64;

Query Match   20.5%; Score 393.5; DB 2; Length 317;
Best Local Similarity   27.5%; Pred. No. 9.1e-26; Indels 11; Gaps 3;
Matches   79; Conservative 87; Mismatches 110;

Qy   69 KAQORLYVASAICLLFMIGEVVGGYLAHSLAVMTDAAHLLTDFASMLISLFSLWMSSRPA 128
          ::| : | :|:.||||.: .:||:||||:|:|:.|:.| : .|:.|:.|:||||||
Db   19 QSKKALWITLVLTVFFTVVEIIGGVLSNSLALLSDSAHMASDVLALGLSMVALYMATRPP 78

Qy   129 TKTMNFGWQRAEILGALVSVLSIWVTVGVLYLYLAVERLISGDYEIDGGTMLITSGCAVAV 188
          ||:| || |:|||||:|||:||||||.|||||:| | :: || |:||||||||| ||
Db   79 NRRFTPGFLRFEIIASPLNGLTLAVIAWILWEGIERFLHPE-PIQFRLMLGLIAAIGLIV 137

Qy   189 NIIMGLTLHQSGHGHSHGTTNQQEENPSVRAAFIHVIGDFMQSMGVLVAAYILYFKPEYK 248
          | ::| :||||:|:||||||||||||||||||||||||||| ::|||||: ||:|:
Db   138 NLTLIVLSRS------TKEEDNLNVQSALWHFIGDLISSGVIISALLIYF-TGWT 187

Qy   249 YVDPICIFVFSILVLIGTTLTLITLIRDVILVLMEGTPKGVDFTAVRDLLLSVEGVEALHSLHI 308
          :|||:| || : ||: :|| ||:|||: :: || ::|||| |||: :||||:|||| |:::
Db   188 IFPDIISLVIAAIIFTGGGAKIMRESYILIMEAVPDEFDLEQIRADIRNIEGVEDVHDMHL 247

Qy   309 WALTVAQPVLSVHIAIAQNTDAQAVLKTASSRLQGFKFHFHVYTIQIE 355
          | || |||:|||||||| ||||:|:|||||||| | | | | :
Db   248 WAISTDHYSLSAHVFVNEHIQPLCVILAVNEMLKEKYGIEHSTIQVE 294

RESULT 32
ID   Q99SB4   PRELIMINARY;   PRT;   325 AA.
AC   Q99SB4;
DT   01-JUN-2001 (TrEMBLrel. 17, Created)
DT   01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
DT   01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE   CzrB protein (Cation-efflux system membrane protein homolog).
GN   CZRB OR SAV2146 OR SA1948.
OS   Staphylococcus aureus (strain Mu50 / ATCC 700699), and
OS   Staphylococcus aureus (strain N315).
OC   Bacteria; Firmicutes; Bacillus/Clostridium group; Bacillales;

OC   Staphylococcus.
OX   NCBI_TaxID=158878, 158879;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   SPECIES=S.aureus (strain Mu50), and S.aureus (strain N315);
RX   MEDLINE=21311952; PubMed=11418146;
RA   Kuroda M., Ohta T., Uchiyama I., Baba T., Yuzawa H., Kobayashi I.,
RA   Cui L., Oguchi A., Aoki K.-I., Nagai Y., Lian J.-Q., Ito T.,
RA   Kanamori M., Matsumaru H., Maruyama A., Murakami H., Hosoyama A.,
RA   Mizutani-Ui Y., Takahashi N.K., Sawano T., Inoue R.-I., Kaito C.,
RA   Sekimizu K., Hirakawa H., Kuhara S., Goto S., Yabuzaki J.,
RA   Kanehisa M., Yamashita A., Oshima K., Furuya K., Yoshino C., Shiba T.,
RA   Hattori M., Ogasawara N., Hayashi H., Hiramatsu K.;
RT   "Whole genome sequencing of meticillin-resistant Staphylococcus
RT   aureus.";
RL   Lancet 357:1225-1240(2001).
DR   EMBL; AP003364; BAB58308.1; -.
DR   EMBL; AP003136; BAB43232.1; -.
DR   InterPro; IPR002524; Cation_efflux.
DR   Pfam; PF01545; Cation_efflux; 1.
DR   TIGRFAMs; TIGR01297; CDF; 1.
KW   Complete proteome.
SQ   SEQUENCE   325 AA;   36129 MW;   C81B33D831689817 CRC64;

Query Match   20.3%; Score 391; DB 16; Length 325;
Best Local Similarity   28.9%; Pred. No. 1.5e-25; Indels 28; Gaps 7;
Matches   95; Conservative 73; Mismatches 133;

Qy   47 AQSNHHCHAQKGPDSHCDPKKGRAQRQLYVASAICLLFMIGEVVGGYLAHSLAVMTDAAH 106
          : | | :   : :. :|||||.:|::|:|||||:|.|||:||||||||||:|:|:|||
Db   2 SHSHHHDHMH---SHVTTNNKKV--LFISFLIGLIYMFIEIIGGLLANSLALLSDGIH 54

Qy   107 LLTDFASMLISLFSLWMSSRPATKTMNFGWQRAEILGALVSVLSIWVTVGVLYLYLAVERL 166
          :|||||||||||||||||||||:|:| || |:|||||:|||:|||||||.|||||:| |
Db   55 MFSDFFSLGVALVAFIYAEKNATTTKFGYKRFEVLAALFNGVTLFVISILLIVFEAIKRF 114

Qy   167 ISGDYEIDGGTMLITSGCAVAVNIIMGLTLHQSGHGHSHGTTNQQEENPSVRAAFIHVIG 226
          ::: |: |:|||||| | | || : | :|||:|:|||||||||||||||||||||||
Db   115 FVPS-EVQSKEMLIISIGLIVNIVAAFMFKGG------DTSHNLNMRGAFLHVIG 164

Qy   227 DFMQSMGVLVAAYILYFKPEYKYVDPICIFVFSILVLGTTLTLITLIRDVILVLMEGTPKGVD 286
          ||:|||||||::| |:|: |:|||:| || : ||: :|| ||:|||: :: ||| ::|||
Db   165 DLLGSVGAITAA-ILIWAFGWTIADPIASILVSVIILKSAWGITKSSINILMEGTPSDVD 223

Qy   287 FTAVRDLLLSVEGVEALHSLHIWALTVAQPVLSVHIAIAQNTDAQAVLKTASSRLQGFKFH 346
          |||:|| :||||:|||| |:: | || |||:|||||||| ||||:|:||||||||| | |
Db   224 IDEVTTTIKKDSRIQSVHDCHVWTISNDMNALSCHVVV---DHTLTMKECELLLENIEH 279

Qy   347 -----FHVTVIQIEDYSEDMKDCQACQG 369
          | | ::|: : |:|
Db   280 DLLHLNIHIHMTIQLETPNHKHDESIICSG 308

RESULT 33
ID   Q92NF5   PRELIMINARY;   PRT;   325 AA.
AC   Q92NF5;
DT   01-MAY-1999 (TrEMBLrel. 10, Created)
DT   01-MAY-1999 (TrEMBLrel. 10, Last sequence update)
DT   01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE   CzrB protein.
GN   CZRB.
OS   Staphylococcus aureus.
OC   Bacteria; Firmicutes; Bacillus/Clostridium group; Bacillales;
OC   Staphylococcus.
OX   NCBI_TaxID=1280;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   STRAIN=912;
RA   Kuroda M., Hayashi H., Ohta T.;
RT   "Chromosome-determined zinc responsible operon czr in Staphylococcus
RT   aureus strain 912.";
RL   Microbiol. Immunol. 0:0-0(1999).

```
DR   EMBL; AB016431; BAA36686.1; -.
DR   InterPro; IPR002524; Cation_efflux.
DR   Pfam; PF01545; Cation_efflux; 1.
DR   TIGRFAMs; TIGR01297; CDF; 1.
SQ   SEQUENCE   325 AA;  36130 MW;  1C5DD3DFD6DD15A2 CRC64;

Query Match             20.3%;  Score 390;  DB 2;  Length 325;
Best Local Similarity   29.3%;  Pred. No. 1.9e-25;
Matches   97; Conservative   73; Mismatches  129; Indels  32; Gaps   8;

Qy     47 AQSNHHCHAQKGPDSH--CDPKKGKAQRQLYASAICLLFMIGEVGGIYLAHSLAVMTDA 104
          |:  :| |  | :||||  | | ||||| |::::|:: | | |||::|||||||:::|
Db      2 SHSHHHDHMH----SHVTTDNKK----VLFISFLIIGLYMFIEIIGGLLANSLALLSDG 52

Qy    105 AHLLTDFASMLISLFSLWMSSRPATKTMNFGWMQRAEILGALVSVLSIWVVTGVLVYLAVE 164
          |::| |||||||||||||||||||||||||||||||| |||:|||:::|:||| :||||
Db     53 IHMFSDTFSLGVALVAFIYAEKNATTTKFTFGYKRFEVLAALFNGVTLVISILIVFEAIK 112

Qy    165 RLISGDYEIDGGTMLITSGCAVAVNIIMGLTLHQSGHGHSHGTTNQQEENPSVRAAFIHV 224
          |:    :   ||||:||||||||||||||||||||:||:||||||:|:||||||||||||
Db    113 RFFVPS-EVQSKEMLISIGLIVNIVAFFMFKGG--------DTSHNLNMRGAFLHV 162

Qy    225 IGDFMQSMGVLVAAYILYFKPEYKVDPICTFVFSILVLGTTLTILRDVILVLMEGTPKG 284
          ||| : |:||| | :|| ||:|:|:||| : :|:|::|| | ::||||||||||||||:|
Db    163 IGDLLGSVGAITAA-ILIWAFGWTIADPIASILVSVIILKSAWGITKSSINILMEGTPSD 221

Qy    285 VDFTAVRDLLLSVEGVEALHSLHIWALTVAQPVLSVHIAIAQNTDAQAVLKTASSRLGGK 344
          ||   :  : :  : || :| | |  ||:::: |||| : :::    :| ::||: | :|
Db    222 VDIDEVITTIKKDSRIQSVHDCHVWTISNDMNALSCHVVV----DHVILTMKECELLENI 277

Qy    345 FH------FHTVTIQIEDYSEDMKDCQACQG 369
          ||      ||:||||:|  |::  |:: :||
Db    278 EHDLLHLNIHHHMTIQLETPNHKHDESIICSG 308
```

```
RESULT 34
P71023
ID   P71023;     PRELIMINARY;    PRT;   295 AA.
AC   P71023;
DT   01-FEB-1997 (TrEMBLrel. 02, Created)
DT   01-FEB-1997 (TrEMBLrel. 02, Last sequence update)
DT   01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE   CzcD (Fragment).
GN   CzcD.
OS   Bacillus subtilis.
OC   Bacteria; Firmicutes; Bacillus/Clostridium group; Bacillales;
OC   Bacillaceae; Bacillus.
OX   NCBI_TaxID=1423;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=JH642;
RX   MEDLINE=97254454; PubMed=9099864;
RA   Sturr M.G., Ablooglu A.J., Krulwich T.A.;
RT   "A Bacillus subtilis locus encoding several gene products affecting
RT   transport of cations.";
RL   Gene 188:91-94(1997).
DR   EMBL; U62055; AAB53029.1; -.
DR   InterPro; IPR002524; Cation_efflux.
DR   Pfam; PF01545; Cation_efflux; 1.
DR   TIGRFAMs; TIGR01297; CDF; 1.
FT   NON_TER       1      1
SQ   SEQUENCE   295 AA;  32404 MW;  B5952B7CD517C7BE CRC64;

Query Match             20.2%;  Score 388.5;  DB 2;  Length 295;
Best Local Similarity   28.8%;  Pred. No. 2.2e-25;
Matches   83; Conservative   81; Mismatches 109; Indels  15; Gaps   5;

Qy     84 FMIGEVVGGYLAHSLAVMTDAAHLLTDFASMLISLFSLWMSSRPATKTMNFGWMQRAEIIG 143
          :|| | ||:|: |||||:|| |||:|  :|| |:|:: || :::| |||:|||||:|||
Db      8 YMIEAIGGFLINSLALLSDAGHMLSDSISLMVALIAFTLAEKKANHNKTFGYKRFEILA 67

Qy    144 ALVSVLSIWVVTGVLVLYLAVERLISGDYEIDGGTMLITSGCAVAVNIIMGLTLHQSGHGH 203
          |:|  | |:: :|:| | | |||| || || ||:|||||||||||||||||||||||||||
Db     68 AVINGAALLILSLIIYEAIER-FSNPPKYATTGMLTISIGLIVVNLLIVAWIMSGG--- 123

Qy    204 SHGTTNQQEENPSVRAAFIHVIGDFMQSMGVLVAAYILYFKPEYKXVDPICTFVFSILVL 263
                  :||  :| |||:||:|  :| |:||  : | ||  || | |  : : :|: ||
Db    124 ------DTKNLNIRGAYLHVISDMLGSVGAILAILIIFF-GWGWANPLASIIVAILVL 176

Qy    264 GTTLTILRDVILVLMEGTPKGVDFTAVRDLLLSVEGVEALHSLHIWALTVAQPVLSVHIA 323
          || ::| :|:| ||||||| :|  :::::  : |:  :: | ::|  |::::  :|| |||
Db    177 RSGYNVTKDSIHILMEGTPNIDVSDIIRTIEGTEGIQNIHIDLHIWSITSGLNALSCHAV 236

Qy    324 IAQN--TDAQAVLKTASSRLQGKFHFHTVTIQIEDYSEDMKDCQACQG 368
          : :   ::  ::||:::: | ||  :|:||:| : :||: ::::|||
Db    237 VDDQLTISESENILRKIEHELEHKGITH-VTIQMETEAHNHDNAILCQ 283
```

```
RESULT 35
Q9ABC7
ID   Q9ABC7;     PRELIMINARY;    PRT;   361 AA.
AC   Q9ABC7;
DT   01-JUN-2001 (TrEMBLrel. 17, Created)
DT   01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
DT   01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE   Cation efflux family protein.
GN   CC0303.
OS   Caulobacter crescentus.
OC   Bacteria; Proteobacteria; alpha subdivision; Caulobacter group;
OC   Caulobacter.
OX   NCBI_TaxID=155892;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=ATCC 19089 / CB15;
RX   MEDLINE=21173698; PubMed=11259647;
RA   Nierman W.C., Feldblyum T.V., Laub M.T., Paulsen I.T., Nelson K.E.,
RA   Eisen J., Heidelberg J.F., Alley M.R.K., Ohta N., Maddock J.R.,
RA   Potocka I., Nelson W.C., Newton A., Stephens C., Phadke N.D., Ely B.,
RA   DeBoy R.T., Dodson R.J., Durkin A.S., Gwinn M.L., Haft D.H.,
RA   Kolonay J.F., Smit J., Craven M.B., Khouri H., Shetty J., Berry K.,
RA   Utterback T., Tran K., Wolf A., Vamathevan J., Ermolaeva M., White O.,
RA   Salzberg S.L., Venter J.C., Shapiro L., Fraser C.M.;
RT   "Complete genome sequence of Caulobacter crescentus.";
RL   Proc. Natl. Acad. Sci. U.S.A. 98:4136-4141(2001).
DR   EMBL; AE005704; AAK22290.1; -.
DR   TIGR; CC0303; -.
DR   InterPro; IPR002524; Cation_efflux.
DR   InterPro; IPR002395; Kininogen.
DR   Pfam; PF01545; Cation_efflux; 1.
DR   PRINTS; PR00334; KININOGEN.
DR   TIGRFAMs; TIGR01297; CDF; 1.
KW   Complete proteome.
SQ   SEQUENCE   361 AA;  38180 MW;  1A4F7F0A7C62EEB0 CRC64;

Query Match             19.5%;  Score 374.5;  DB 16;  Length 361;
Best Local Similarity   28.3%;  Pred. No. 4.6e-24;
Matches   92; Conservative   70; Mismatches 140; Indels  23; Gaps   8;

Qy     50 NHHCHAQKGPDSHCDPKKGKAQRQLYASAICLLFMIGEVVGGIYLAHSLAVMTDAAHLLT 109
          :|| | |   | | ||| ||:   |  | | | | :|| |:| |: |||||:|| |||:|
Db     56 HHHGHGHGHHHHAPKDFGRA---FAIGTALNLGFVIVEATAGILTHSLALLADAGHNLS 112

Qy    110 DFASMLISLFSLWMSSRPATKTMNFGWMQRAEILGALVSVLSIWVVTGVLVYLAVERLISG 169
           : |:|  ||: ::::|||: : : ||:: |:|  |||: | |:: :|:| | | |||: |
Db    113 DVLGLLAWGATVLAKRAPSARRTVGLRKGTILASLGNAALLLAVGAIAWEGVRRFAAP 172

Qy    170 DYEIDGGTMLITSGCAVAVNIIMGLTLHQSGHGHSHGTTNQQEENPSVRAAFIHVIGDFM 229
          : |: | ||::: |  |::| : : |||  |   |::    :||  :| |||:||: :|
Db    173 E-PVQTGPVMIVAAIGIVINPATALMFMKG------SKEDLNVRGAFLHMAAADA 220

Qy    230 QSMGVLVAAYILYFKPEYKXVDPICTFVF-SILVLGTTLTILRDVILVLMEGTPKGVDFT 288
           |:|::: |  : | :  :| |||: ::| |:|:||: : || |::::: | : |:|:
Db    221 VSAGVVIAALAMTF-TGWWMLDPVVSLVIVAVIVLG-TWGLLRDSIDLALDATPRGIDTQ 278

Qy    289 AVRDLLLSVEGVEALHSLHIWALTVAQPVLSVHIAIAQNTDAQAVLKTASSRLQGKFHFH 348
          ::::  : |:  ||:| : |  |::::::|||| : :::    :|::::|:  | ||
Db    ...
```

Db   279 KVRDWLAARPGVSEVHDLHIWAMSTTETALTFAHVVRQLDADHDQFLHDACAELASRFNIG 338

Qy   349 TVTIQIEDYSEDMKDCOACQ-GPSD 372
                |:| || |:| |:|
Db   339 HVTIQV----ESGHGAHACRLAPAD 359

RESULT 36
ID   Q9I6A3   PRELIMINARY;   PRT;   299 AA.
AC   Q9I6A3;
DT   01-MAR-2001 (TrEMBLrel. 16, Created)
DT   01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DT   01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE   Probable cation efflux system protein.
GN   PA0397.
OS   Pseudomonas aeruginosa.
OC   Bacteria; Proteobacteria; gamma subdivision; Pseudomonadaceae;
OC   Pseudomonas.
OX   NCBI_TaxID=287;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=ATCC 15692 / PAO1;
RX   MEDLINE=20437337; PubMed=10984043;
RA   Stover C.K., Pham X.-Q.T., Erwin A.L., Mizoguchi S.D., Warrener P.,
RA   Hickey M.J., Brinkman F.S.L., Hufnagle W.O., Kowalik D.J., Lagrou M.,
RA   Garber R.L., Goltry L., Tolentino E., Westbrock-Wadman S., Yuan Y.,
RA   Brody L.L., Coulter S.N., Folger K.R., Kas A., Larbig K., Lim R.M.,
RA   Smith K.A., Spencer D.H., Wong G.K.-S., Wu Z., Paulsen I.T.,
RA   Reizer J., Saier M.H., Hancock R.E.W., Lory S., Olson M.V.;
RT   "Complete genome sequence of Pseudomonas aeruginosa PAO1, an
RT   opportunistic pathogen.";
RL   Nature 406:959-964(2000).
DR   EMBL; AE004477; AAG03786.1; -.
DR   InterPro; IPR002524; Cation_efflux.
DR   Pfam; PF01545; Cation_efflux; 1.
DR   TIGRFAMs; TIGR01297; CDF; 1.
KW   Complete proteome.
SQ   SEQUENCE   299 AA;   32373 MW;   C35EEDDFABD04B10 CRC64;

Query Match   19.3%; Score 372; DB 16; Length 299;
Best Local Similarity 32.2%; Pred. No. 5.9e-24;
Matches   88; Conservative   62; Mismatches 109; Indels   14; Gaps   5;

Qy   84 FMIGEVVGGYLAHSLAVMTDAAHLLTDFASMLISLFSLWMSSRPAIKTMNFGWQRAEILG 143
               |:|| | |||| || | |||||:|||||:||||:||||:: | | ||| | | ::||
Db   27 FLVAEVVGGILTGSLALISDAAHMLTDAVALAIALAAINIARRPTNDRLTYGYHREFILA 86

Qy   144 ALVSVLSIWVTVGVLVLAVERLISGDYEIDGGTMLITSGCAVAVNIIMGLTLHQSGHGH 203
               ::| |::| | || |:|||||| ||||||:|| ||||| : ||: | ||||
Db   87 AAFNAFLLFGVAFYILYAAYERL-NQPAEIQSVGMLVIAVLGLLVN--LASMRLLAPAQGN 144

RESULT 37
ID   O85143   PRELIMINARY;   PRT;   326 AA.
AC   O85143;
DT   01-NOV-1998 (TrEMBLrel. 08, Created)
DT   01-NOV-1998 (TrEMBLrel. 08, Last sequence update)
DT   01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE   Transport protein.
RC   STRAIN=N.A.
RP   SEQUENCE FROM N.A.
RN   [1]
OX   NCBI_TaxID=7227;
OC   Ephydroidea; Drosophilidae; Drosophila.
OC   Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC   Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OS   Drosophila melanogaster (Fruit fly).
GN   CG17723.
DE   CG17723 protein (LD22804P).
DT   01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DT   01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT   01-MAY-2000 (TrEMBLrel. 13, Created)
AC   Q9VZR4;
ID   Q9VZR4   PRELIMINARY;   PRT;   449 AA.
RESULT 38

Qy   347 ---FHTVTIQIEDYSEDMKDCQACOQ 369
                  ||||:|:| |
Db   284 HLNIHHMSIQLETPNHKHDESIICSG 309

Qy   290 VRDLLLSVEGVEALHSLHIWALTVAQPVLSVHIAIAQNTDAQAVLKTASSRLQGKFH--- 346
               ::::| ||| ::| :| :::::| ||| || :|:: ||:|:|:|| |
Db   228 VITITKDSRIQSVHDCHVWTISNDMNALSCHVVV----DHTLTMKECELLENIEHDLL 283

Qy   230 QSMGVLVAAYILYFKPEYKYVDPICTFVFSILVL 263
               :| ||: :| ||::|| | |::| :| :| |:
Db   119 S-EVQSKEMLIISIIGLLIVNIVAFFRKGG----- 168

Qy   170 DYEIDGGTMLITSGCAVAVNIIMGLTLHQSGHGHGHSHGTTNQQEENPSVRAAFIHVIGDM 229
               | :|||||:|| | :| ||:| ||| | | ::||| :||: ::: ||| ||| |
Db   59 DTFSLGVALVAFIYAEKNATTIKFGYKRFEVLAALFNGVTLFVISILIVFEAITRFVP 118

Qy   110 DFASMLISLFSLWMSSRPAIKTMNFGWQRAEILGALVSVSLSIWVTVGVLVLYLAVERLISG 169
               ||:|||:||||:||||| |::|| ||:|:| ||:|: | ::| ||| :| |:| ||||:|
Db   6 HHIDHMH----SHVTTNN--KEVLFISPLIIGLYMFIEIIGGLLANSLALLRDGIHMSS 58

Qy   50 NHHCHAQKGPDSHCDPKKGKAQRQLIVASAICLLFMIGEVVGGYLAHSLAVMTDAAHLLT 109
               | | |:| :| ||| | ::| | :| | | |||||||||||||||| |||| | |
Db   170 DYEIDGGTMLITSGCAVAVNIIMGLTLHQSGHGHGHSHGTTNQQEENPSVRAAFIHVIGDM 229

SQ   SEQUENCE   326 AA;   36233 MW;   47311E0C1FFDE4FC CRC64;
TIGRFAMs; TIGR01297; CDF; 1.
Pfam; PF01545; Cation_efflux; 1.
InterPro; IPR002524; Cation_efflux.
EMBL; AF044951; AAC32485.1; -.
J. Bacteriol. 180:4024-4029(1998).
"Molecular characterization of a chromosomal determinant conferring
resistance to zinc and cobalt ions in staphylococcus aureus.";
Xiong A., Jayaswal R.K.;
MEDLINE=98361887; PubMed=9667746;
STRAIN=RN450;
SEQUENCE FROM N.A.
[1]
NCBI_TaxID=1280;
Staphylococcus.
Staphylococcus;
Bacteria; Firmicutes; Bacillus/Clostridium group; Bacillales;
Staphylococcus aureus.

Query Match   19.2%; Score 369; DB 2; Length 326;
Best Local Similarity 28.2%; Pred. No. 1.2e-23;
Matches   92; Conservative   71; Mismatches 135; Indels   28; Gaps   7;

RESULT 36 ... (continued)

DE   YDPALVDAEALLGTVKALLHDRYEIEHSTLQLE 286
                    |:|:| | ||:| | ||| :|:|
Qy   254 I-AQNTDAQAVLKTASSRLQGKFHFHFTVTIQIE 355
                  | |:|:|:| :|| |:|| :|||::|:
Qy   324 PRTWILLRESLHVLLEGVPKEIQLAELREALLGIPGVTGLHDLHVWSITSGKISLTSHLV 253
                ||:|| :| || || ::| : :| | || | : | ||| || | | | ::|:
Db   194 GTTLTILRDVLILVLMEGTPKGVDFTAVRDLLLSVEGVEALHSLHIWALTVAQPVLSVHIA 323
Db   264

Qy   204 SHGTTNQQEENPSVRAAFIHVIGDFMQSMGVLVAAYILYFKPEYKYVDPICTFVFSILVL 263
                    ::|| |:||:|: || | | | :: | :|| | |
Db   145 S-------LNVKGAYLEVWSDMLGSLGVIVAAIVIRF-TGWAAWVDSLVAVLIGFWVL 193

RA Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
RA Borkova D., Botchan M.R., Bouck J., Brokstein P., Brottier P.,
RA Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA Durbin K.J., Evangelista C.C., Ferraz C., Ferriera S., Fleischmann W.,
RA Fosler C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,
RA Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacleb J.M.,
RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
RA Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA Wang Z.-Y., Wassarman D.A., Weinstock G.M., Weissenbach J.,
RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
RL "The genome sequence of Drosophila melanogaster.";
RL Science 287:2185-2195(2000).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=BERKELEY;
RA Stapleton M., Brokstein P., Hong L., Agbayani A., Carlson J.,
RA Champe M., Chavez C., Dorsett V., Farfan D., Frise E., George R.,
RA Gonzalez M., Guarin H., Li P., Liao G., Miranda A., Mungall C.J.,
RA Nunoo J., Pacleb J., Paragas V., Park S., Phouanenavong S., Wan K.,
RA Yu C., Lewis S.E., Rubin G.M., Celniker S.;
RL Submitted (OCT-2001) to the EMBL/GenBank/DDBJ databases.
DR EMBL; AE003477; AAF47754.1; -.
DR EMBL; AY058521; AAL13750.1; -.
DR FlyBase; FBgn0035432; CG17723.
DR InterPro; IPR002524; Cation_efflux.
DR Pfam; PF01545; Cation_efflux; 1.
DR TIGRFAMs; TIGR01297; CDF; 1.
SQ SEQUENCE   449 AA; 49032 MW; 216F49852773F44B CRC64;

Query Match   19.0%; Score 366; DB 5; Length 449;
Best Local Similarity   28.9%; Pred. No. 3.3e-23;
Matches 103; Conservative 74; Mismatches 117; Indels 62; Gaps 14;

Qy   68 GKAQRQLYVASAICLLFMIGEVGGYLAHSLAVMTDAAHLLTDFASMLISLFSLWMSSRP 127
             || |||| |:|:| | |: |||| | |||| |:::|: || :|: :|:| |
Db   6 GKKCRLLSMMWLTAFFFV-EIIVGYVTNSMALVADSFHMLGDIAALVISFLSVKMSPKK 64

Qy 128 ATKTMNFGWQRAEILGALVSVLSIWVVTGVLVYLAVERLISGDYEIDGGTMLITSGCAVA 187
             |||||||||:||||:|:||| |::|:|| :|| | || || :| | |: |:||:||:
Db   65 WSKN-TFGWARAEVLGALVNAVFLVALCFSITEACKRFIEEPIHEPELLIVGALGLL 123

Qy 188 VNIIMGLTL---HQSGHGHSHG---TIN---------   QQEENP-- 215
             |:|| |:| | | | || | |:| |||||
Db 124 VNVI-GLCLLYEHGGHGHSHGGGLTRNHSRLTELANMDEGEDEQNDFAYEKOKEKAPVK 182

Qy 216 ----------SVRAAFIHVIGDFMQSMGVLVAAVILYFKPEYK----YVDPICTFVPS 259
                       :|| |||| | |:| |: :|| ||||:: || :|: |: |
Db 183 KSSHGHSHDPGQMNMRGAFLHVLSDALGSIIVVISAVVVW-KTEWKIRYYMDPALSIVLV 241

Qy 260 ILVLGTTLTILILRDVILVLMEGTPKGVDFTAVRDLLL-SVEGVEALHSLHIWALTVAQPVL 318
         ||:|:| ||: |: :: |: || :| |||| | | | ||||| |||| | || ||| |
Db 242 VILLHSWPLLRESALILLQTVPTHIQVDAIQKRLLEKVDGVLAVHEFHVWQLAGDRIIA 301

Qy 319 SVHAIAQNTDAQAVLKTASSRLQGKFH---FHTVTIQIEDYSEDMKDCQACQGPS 371
         :| :| |||||| |||: :|:| || | |||| |||||| ||| ||| ||:
Db 302 SAHIRCRNLSEYMKI---AEKVKEFFHNEGIHSTTIQPE-FSE-IEGCNMSDGTS 351

RESULT 39
Q98H85
ID   Q98H85   PRELIMINARY;   PRT;   365 AA.
AC   Q98H85;
DT   01-OCT-2001 (TrEMBLrel. 18, Created)
DT   01-OCT-2001 (TrEMBLrel. 18, Last sequence update)
DT   01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE   Cation efflux system protein.
GN   MLL2984.
OS   Rhizobium loti (Mesorhizobium loti).
OC   Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group;
OC   Phyllobacteriaceae; Mesorhizobium.
OX   NCBI_TaxID=381;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=MAFF303099;
RX   MEDLINE=21082930; PubMed=11214968;
RA   Kaneko T., Nakamura Y., Sato S., Asamizu E., Kato T., Sasamoto S.,
RA   Watanabe A., Idesawa K., Ishikawa A., Kawashima K., Kimura T.,
RA   Kishida Y., Kiyokawa C., Kohara M., Matsumoto M., Matsuno A.,
RA   Mochizuki Y., Nakayama S., Nakamura Y., Shimpo S., Sugimoto M.,
RA   Takeuchi C., Yamada M., Tabata S.;
RT   "Complete genome structure of the nitrogen-fixing symbiotic bacterium
RT   Mesorhizobium loti.";
RL   DNA Res. 7:331-338(2000).
DR   EMBL; AP003000; BAB49981.1; -.
DR   InterPro; IPR002524; Cation_efflux.
DR   Pfam; PF01545; Cation_efflux; 1.
DR   TIGRFAMs; TIGR01297; CDF; 1.
KW   Complete proteome.
SQ   SEQUENCE   365 AA; 38979 MW; A513207601163CEC CRC64;

Query Match   18.7%; Score 360; DB 16; Length 365;
Best Local Similarity   27.2%; Pred. No. 8.2e-23;
Matches 84; Conservative 75; Mismatches 130; Indels 20; Gaps 6;

Qy   52 HCHAOKGPD----SHCDPKKGKAQRLYVASAICLLFMIGEVGGYLAHSLAVMTDAAHL 107
             | | ||| ||: | |||| |||| |:|:| | |: |||| | ||||| |
Db   43 HSHDHSGHDHEGAGHVHGATDK--KRVLIAACLTAGFMVAEALGGLFTGSLALLADAGHM 100

Qy 108 LTDFASMLISLFSLWMSSRPATKTMNFGWQRAEILGALVSVLSIWVVTGVLVYLAVERLI 167
         |||||:|:|| :|:|:|:| | :|| ||||:|:|||| :::|:|| :|| | ||:| :
Db 101 LADAIALGLLAWYAFHLAGRPATGQLIYGFGRVKTLVAYTNGIAIFVIALWIVYEAWQRLL 160

Qy 168 SGDYEIDGGTMLITSGCAVAVNIIMGLTLHQSGHGHSHGHGTNQQEENPSVRAAFIHVIGD 227
         :||| | |: |:||:|||:|| |:| |: |:| | ||| | |||| | | || |||| |
Db 161 T-PAPVLGGPMLVVAILGLLVNIGSPLVLH----------GGDRESLNMRGAILHVLGD 208

Qy 228 FMQSMGVLVAAVILYFKPEYKXVDPICTFVPSILVLGTTLTILILRDVILVLMEGTPKGVDF 287
         ||| |:|:||| || || ||| | | :|: |:|:||||: |: :: |: || |: || |
Db 209 LLGSAAAIVAAVVI-LVTGWTPIDPILSVLVSLLILSTAWSLMRAAAHVLLEGVPPSLDR 267

Qy 288 TAV-RDLLLSVEGVEALHSLHIWALTVAQPVLSVHIAIAQNTDAQAVLKTASSRLQGKFH 346
         ||| | || ||||| |||| | ||||||| |||: :| |||||| |||: :|:| ||
Db 268 DLIATDLETTIQGVREVHHMIVWSIDGSSNNATLHACLDEGVDAHRAVSAIKKRLASEHG 327

Qy 347 FHTVTIQIE 355
         |||| |||:
Db 328 ISHAIVEPE 336

RESULT 40
Q06808
ID   Q06808   PRELIMINARY;   PRT;   429 AA.
AC   Q06808;
DT   01-NOV-1996 (TrEMBLrel. 01, Created)
DT   01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT   01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE   Oxidative stress resistance.

GN   ZRC1 OR OSR OR YMR243C.
OS   Saccharomyces cerevisiae (Baker's yeast).
OC   Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC   Saccharomycetales; Saccharomycetaceae; Saccharomyces.
OX   NCBI_TaxID=4932;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=95244001; PubMed=7727035;
RA   Inoue Y.;
RT   "Methylglyoxal and lipid hydroxperoxide as endogenous cytotoxic
RT   molecular species: detoxification and regulation of gene expression in
RT   yeasts.";
RL   Biotechnol. Genet. Eng. Rev. 12:467-508(1994).
RN   [2]
RP   SEQUENCE FROM N.A.
RA   Inoue Y.;
RL   Submitted (MAY-1995) to the EMBL/GenBank/DDBJ databases.
RN   [3]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=93332654; PubMed=8338639;
RA   Inoue Y., Kobayashi S., Yoshikawa K., Tran L., Kimura A.;
RT   "Lipid hydroperoxide-resistance gene in Saccharomyces cerevisiae:
RT   utilization as a selectable marker gene for yeast transformation.";
RL   Biotechnol. Appl. Biochem. 17:305-310(1993).
RN   [4]
RP   SEQUENCE FROM N.A.
RA   Inoue Y., Kobayashi S., Kimura A.;
RL   J. Ferment. Bioeng. 75:327-331(1993).
DR   EMBL; X87634; CAA60962.1; -.
DR   SGD; S0004856; ZRC1.
DR   InterPro; IPR002524; Cation_efflux.
DR   Pfam; PF01545; Cation_efflux; 1.
DR   TIGRFAMs; TIGR01297; CDF; 1.
SO   SEQUENCE   429 AA; 47075 MW; B41C79557D99980E9 CRC64;

Query Match   18.6%; Score 357; DB 3; Length 429;
Best Local Similarity   25.2%;
Matches   97; Conservative   80; Mismatches 108; Indels 100; Gaps 13;

Qy   72 ROLYVASAICL--LFMIGEVVGGYLAHSLAVWTDAAHLLTDFASMLISLFSLWMS-SRPA 128
          ::|:|||:| ::::|:|::|||||||||::||:::|||:::|||::::::: :|
Db   5 KELRIISLLTLDTVFFLLEITIGYMSHSLALIADSFHMLNDIISLLVALWAVDVAKNRGP 64

Qy   129 TKTMNFGWQRAEILGALVSVLSIWVVTGVLVYLAVERLISGDYEIDGGTMLITSGCAVAV 188
          ::|:|||||||||||||:::: :: |||||||||||||||:::|| |||:|||:|:
Db   65 DAKYTYGWKRAEILGALINAVFLIALCFSIMEEALQRLIE-PQEIQNPRLVIYGVAGLI 123

Qy   189 NIIMGLTL-HOSG----HGHSHGTIN------------------------------- 209
          ::||:| | | | | :||:
Db   124 SNVVGLFLFHDHGSDSLHSHSHGSVESGNNDLDIESNATHSHSHASLPNDNLAIDEDAIS 183

Qy   210 ----------------QOEENP------------------SVRAAFIHVIGDF 228
                          :: |:|::||:|| ||
Db   184 SPGPSGQIGEVLPQSVVNRLSNESQPLLNHDDHDHSHESKKPGHRSLNMHGVFLHVLGDA 243

Qy   229 MQSMGVLVAAVILYKPEYK----YVDPICTFVFSILVLGTTLTILRDVILVLMEGTPKGV 285
          :|:::||| |||| | | || |:|||:||||:|:::|:|:||:|:||| |:|
Db   244 LGNIGVIAAALFIW-KTEYSWRVYSDPIVSLITIIIFSSALPLSRRASRILLQATPSTI 302

Qy   286 DFTAVRDLLLSVEGVEALHSLHIWALTVAQPVLSVHIAIAQNTDAQAVLKTASSRLQGK- 344
          :::::|:| |: :||:|:: ||:| ::::|:   :|: | ::|::::||::||:|
Db   303 SADQIQREILAVPGVIAVHDFHVWNLTESIYIASHVQI---DCAPDKFMSSAKLIRKI 358

Qy   345 FH---FHTVTIQIE----DYSEDMK 362
          || ||:|:|:| | :|:|:
Db   359 FHQHGIHSATVQPEFVSGDVNEDIR 383

Search completed: March 18, 2003, 15:42:14
Job time : 44 secs

GenCore version 5.1.4_p5_4578
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: March 18, 2003, 15:36:44 ; Search time 11 seconds
(without alignments)
1402.654 Million cell updates/sec

Title: US-09-691-219-2
Perfect score: 1923
Sequence: 1 MEAKEKQHLLDAFPAIRSYT.........QIEDYSEDMKDCQACQGPSD 372

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 112892 seqs, 41476328 residues

Total number of hits satisfying chosen parameters: 112892

Post-processing: Minimum Match 0%
Maximum Match 100%

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Database : SwissProt.40.*
Maximum Match 100%
listing first 120 summaries

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	Match DB	ID	Accession	Description
1	1493	77.6	359	1	ZNT2_RAT	Q62941	rattus norv
2	1006.5	52.3	388	1	ZNT3_MOUSE	P97441	mus musculu
3	972.5	50.6	388	1	ZNT3_HUMAN	Q99726	homo sapien
4	722.5	37.6	429	1	ZNT4_HUMAN	O14863	homo sapien
5	716	37.2	430	1	ZNT4_MOUSE	O35149	mus musculu
6	711	37.0	430	1	ZNT4_RAT	O55174	rattus norv
7	434	22.6	312	1	ZITB_SALTY	Q82gt3	salmonella
8	431	22.4	312	1	ZITB_SALTI	Q82db6	salmonella
9	431	22.4	313	1	ZITB_ECOLI	Q82bb6	escherichia
10	427	22.2	311	1	ZITB_ECO57	P75757	escherichia
11	388	20.2	316	1	ZCCD_ALCEU	O2218	alcaligenes
12	387	20.1	316	1	ZCCD_ALCSP	O8x400	alcaligenes
13	380.5	19.8	316	1	ZITB_YERPE	P13512	yersinia pe
14	357	18.6	442	1	ZRC1_YEAST	P94178	saccharomyc
15	337	17.5	439	1	COT1_YEAST	O82gy6	saccharomyc
16	335	17.4	507	1	ZNT1_RAT	P20107	rattus norv
17	333	17.3	503	1	ZNT1_MOUSE	P32798	mus musculu
18	173	9.0	310	1	YC63_SYNY3	O62720	synechocyst
19	153	8.0	108	1	YGLB_BACST	O60738	bacillus st
20	143	7.4	290	1	YEAB_BACSU	P74068	bacillus su
21	132.5	6.9	510	1	YM43_YEAST	P30540	saccharomyc
22	130.5	6.8	300	1	YIIP_ECOLI	P46348	escherichia
23	130	6.8	283	1	Y449_METJA	O02218	methanococc
24	107	5.6	251	1	YQR5_CAEEL	O57891	caenorhabdi
25	104.5	5.4	515	1	KE4L_CAEEL	O09308	caenorhabdi
26	102.5	5.3	379	1	CYB_PHOHI	O35468	phoca hispi
27	101.5	5.3	681	1	NU5C_SYNY3	O35429	synechocyst
28	100.5	5.2	379	1	CYB_ORNAN	Q9xcq7	ornithorhyn
29	100	5.2	404	1	YGJK_CAEEL	O36661	caenorhabdi
30	99	5.2	378	1	CYB_DROYA	O9xuc4	drosophila
31	98.5	5.1	379	1	CYB_CYSCR	P07704	cystophora
32	97.5	5.1	381	1	CYB_PHATA	O35673	phascogale
33	97	5.0	50	1	ZNT4_BOVIN	O9ttf3	bos taurus
34	97	5.0	482	1	ARCD_PSEAE	P18275	pseudomonas
35	96.5	5.0	1021	1	TSCC_HUMAN	P55017	homo sapien
36	96.5	5.0	380	1	CYB_HYLLA	Q95711	hylobates l
37	96	5.0	413	1	MNTH_SALTI	Q8z4x5	salmonella
38	96	5.0	413	1	MNTH_SALTY	Q9rpf4	salmonella
39	96	5.0	605	1	YN8N_YEAST	P53735	saccharomyc
40	95.5	5.0	664	1	NU5C_SYNP2	P31971	synechococc
41	94.5	4.9	233	1	COX3_SYNY3	Q06475	synechocyst
42	94.5	4.9	379	1	CYB_PHOGR	Q35457	phoca groen
43	94.5	4.9	381	1	CYB_ANTFL	O33706	antechinus
44	93.5	4.9	379	1	CYB_PHOFA	O35438	phoca fasci
45	93.5	4.9	380	1	CYB_LATCH	O03176	latimeria c
46	93.5	4.9	1002	1	TSCC_RAT	P55018	rattus norv
47	93.5	4.9	4870	1	RYR3_HUMAN	Q15413	homo sapien
48	93	4.8	378	1	CYB_DROME	P18935	drosophila
49	93	4.8	1094	1	ATHL_RABIT	Q9tv52	oryctolagus
50	92.5	4.8	379	1	CYB_ARCFO	O33697	arctocephal
51	92.5	4.8	381	1	CYB_ANTBE	O63535	antechinus
52	92	4.8	578	1	DSBD_PSESP	O63536	pseudomonas
53	91.5	4.8	381	1	CYB_ANTGO	O63537	antechinus
54	91.5	4.8	381	1	CYB_ANTLE	P92509	antechinus
55	91.5	4.8	381	1	CYB_ANTST	O35409	antechinus
56	91.5	4.8	381	1	CYB_PHACL	P77304	phascogale
57	91.5	4.8	500	1	YDGR_ECOLI	P35409	escherichia
58	90.5	4.7	379	1	CYB_MIRLE	O20433	mirounga le
59	90.5	4.7	381	1	CYB_SMIMA	O35019	sminthopsis
60	90	4.7	949	1	PMA6_ARATH	P41143	arabidopsis
61	90	4.7	372	1	OPRD_HUMAN	O9sh76	homo sapien
62	89.5	4.7	381	1	CYB_HALGR	P38593	halichoerus
63	89.5	4.7	379	1	CYB_LEPWE	P38594	leptonychot
64	89.5	4.7	379	1	CYB_PHOVI	O00530	phoca vitul
65	89.5	4.7	379	1	CYB_TREOR	Q36089	tremarctos
66	89.5	4.7	381	1	CYB_NEOLO	O35157	neophascoga
67	89.5	4.7	381	1	CYB_PSEWO	O35695	pseudantech
68	89.5	4.7	382	1	CYB_DIDMA	P41303	didelphis m
69	89.5	4.7	380	1	CYB_POLOR	O95920	polypterus
70	89	4.6	412	1	MNTH_ECOLI	P77145	escherichia
71	89	4.6	481	1	Y389_CLOPE	P30818	clostridium
72	88.5	4.6	481	1	CYB_ARCGZ	O33688	arctocephal
73	88.5	4.6	379	1	CYB_MONMO	P92657	monodon mon
74	88.5	4.6	379	1	CYB_MOSLE	O47584	moschus leu
75	88.5	4.6	379	1	CYB_MOSMO	O47583	moschus mos
76	88.5	4.6	379	1	CYB_ODOHE	P24960	odocoileus
77	88.5	4.6	379	1	CYB_PHOLR	Q35505	phoca largh
78	88.5	4.6	379	1	CYB_DASSP	O34382	dasyurus sp
79	88.5	4.6	381	1	CYB_MOUSE	P00158	mus musculu
80	88.5	4.6	381	1	CYB_PSEMD	O03543	pseudantech
81	88.5	4.6	381	1	CYB_SMIMU	O35886	sminthopsis
82	88.5	4.6	454	1	FEM2_HUMAN	P49593	homo sapien
83	88.5	4.6	527	1	YEGH_ECOLI	P76389	escherichia
84	88.5	4.6	1205	1	NKC1_MOUSE	P55012	mus musculu
85	88	4.6	155	1	KE4_PIG	Q29175	sus scrofa
86	88	4.6	222	1	YJ73_SYNY3	P73933	synechocyst
87	88	4.6	379	1	CYB_GAZGA	O48336	gazella gaz
88	87.5	4.6	379	1	CYB_DASVI	O34399	dasyurus vi
89	87.5	4.6	381	1	CYB_MACRO	P92671	macropus ro
90	87.5	4.6	395	1	SOTB_ERWCH	O9s319	erwinia chr
91	87	4.6	381	1	CYB_ANTSW	O33865	antechinus
92	87		465	1	ALST_BACSU	O45068	bacillus su
93	87		379	1	CYB_BEAHU	Q47420	beatragus h
94	86.5	4.5	379	1	CYB_DAMPY	O9t7a7	damaliscus
95	86.5	4.5	379	1	CYB_DELLE	O03812	delphinapte
96	86.5	4.5	379	1	CYB_ERIBA	O34409	erignathus
97	86.5	4.5	379	1	CYB_FELCA	P48886	felis silve
98	86.5	4.5	379	1	CYB_URSTH	O36209	ursus thibe
99	86.5	4.5	381	1	CYB_DASBY	O34400	dasyuroides
100	86.5	4.5	381	1	CYB_DASMA	O34300	dasyurus ma
101	86.5	4.5	381	1	CYB_NINRI	O35172	ningaui rid
102	86.5	4.5	381	1	CYB_NINYV	O35196	ningaui yvo
103	86.5	4.5	381	1	CYB_NINYU	O35377	ningaui yvo
104	86.5	4.5	381	1	CYB_PARAP	O35196	parantechin
105	86.5	4.5	381	1	CYB_PARI	O35380	parantechin
106	86.5	4.5	607	1	KAIB_SCHPO	O09898	schizosacch

RESULT 1

```
ID   ZNT2_RAT    STANDARD;    PRT;   359 AA.
AC   Q62941;
DT   30-MAY-2000 (Rel. 39, Created)
DT   30-MAY-2000 (Rel. 39, Last sequence update)
DT   15-JUN-2002 (Rel. 41, Last annotation update)
DE   Zinc transporter 2 (ZnT-2).
GN   SLC30A2 OR ZNT2.
OS   Rattus norvegicus (Rat).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX   NCBI_TaxID=10116;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=96203098; PubMed=8617223;
RA   Palmiter R.D., Cole T.B., Findley S.D.;
RT   "ZnT-2, a mammalian protein that confers resistance to zinc by
RT   facilitating vesicular sequestration.";
RL   EMBO J. 15:1784-1791(1996).
CC   -!- FUNCTION: INVOLVED IN ACCUMULATION OF ZINC IN ENDOSOMAL/LYSOSOMAL
CC       VESICLES.
CC   -!- SUBCELLULAR LOCATION: Integral membrane protein.
CC   -!- LOCALIZED IN THE MEMBRANE OF ENDOSOMAL VESICLES (PROBABLE).
CC   -!- TISSUE SPECIFICITY: INTESTINE, KIDNEY, SEMINAL VESICLES, AND
CC       TESTIS.
CC   -!- SIMILARITY: BELONGS TO THE CATION DIFFUSION FACILITATOR FAMILY OF
CC       TRANSPORTERS (CDF; TC 2.A.4). SLC30A SUBFAMILY.
```

```
DR   EMBL; U50927; AAB02775.1; -.
DR   InterPro; IPR002524; Cation_efflux.
DR   Pfam; PF01545; Cation_efflux; 1.
DR   TIGRFAMs; TIGR01297; CDF; 1.
KW   Zinc transport; Transmembrane; Multigene family; Repeat.
FT   DOMAIN        1     56       CYTOPLASMIC (POTENTIAL).
FT   TRANSMEM     57     77       POTENTIAL.
FT   DOMAIN       78     86       VACUOLAR (POTENTIAL).
FT   TRANSMEM     87    107       POTENTIAL.
FT   DOMAIN      108    123       CYTOPLASMIC (POTENTIAL).
FT   TRANSMEM    124    144       POTENTIAL.
FT   DOMAIN      145    159       VACUOLAR (POTENTIAL).
FT   TRANSMEM    160    180       POTENTIAL.
FT   DOMAIN      181    207       CYTOPLASMIC (POTENTIAL).
FT   TRANSMEM    208    228       POTENTIAL.
FT   DOMAIN      229    236       VACUOLAR (POTENTIAL).
```

```
107  86   4.5   300  1  P34_RICPR      Q9zcc5 rickettsia
108  86   4.5   783  1  YAF3_YEAST     P39719 saccharomyc
109  85.5 4.4   345  1  NQO8_PARDE     P29920 paracoccus
110  85.5 4.4   370  1  Q9mlj8 bungarus fa   Q9mlj8 bungarus fa
111  85.5 4.4   379  1  CYB_GULGU (    P56697 gulo gulo (
112  85.5 4.4   379  1  CYB_HIPNI      Q34760 hippotragus
113  85.5 4.4   381  1  CYB_DASAL      Q34289 dasyurus al
114  85.5 4.4   381  1  CYB_DASCR      Q34302 dasyurcus
115  85.5 4.4   450  1  FEM2_RAT       Q9wvr7 rattus norv
116  85.5 4.4   815  1  AD15_MOUSE     O88839 mus musculu
117  85.5 4.4  1212  1  NKC1_HUMAN     P55011 homo sapien
118  85.5 4.4   374  1  YLD1_CAEEL     O03566 caenorhabdi
119  85   4.4   379  1  CYB_AMMLE      O78781 ammotragus
120  85   4.4   380  1  CYB_PELSU      O79680 pelomedusa
```

ALIGNMENTS

```
FT   TRANSMEM    237    257       POTENTIAL.
FT   DOMAIN      258    291       CYTOPLASMIC (POTENTIAL).
FT   TRANSMEM    292    312       POTENTIAL.
FT   DOMAIN      313    359       VACUOLAR (POTENTIAL).
FT   DOMAIN      183    192       VACUOLAR (POTENTIAL).
SQ   SEQUENCE    359 AA;    39276 MW;   5 X 2 AA APPROXIMATE REPEATS OF H-G.
                256A307021D1D197 CRC64;

Query Match             77.6%;   Score 1493;   DB 1;   Length 359;
Best Local Similarity   80.1%;   Pred. No. 1.4e-103;
Matches  290;  Conservative  35;  Mismatches  29;  Indels  8;  Gaps  2;

Qy    15  AIRSYTGSLWQEGAGWIPLPRPGIDLQAIEIAAQSNHHCHAQKGPDSHCDPKGKAQRQL   74
          ||||:||:.:||||:|.|||||||||||||||||||||:||:|:|
Db     2  ASRSFFGALWKSEASRI----PPVNLPSVELAVQSNHYCHAQKDpSGSHPNSEKQRARRKL  57

Qy    75  YVASAICLLFMIGEVVGGYLAHSLAVMTDAAHLLTDFASMLISLFSLWMSSRPATKTMNF  134
          |||||||:.|||||:|:|||||:|||:|||||||||||||||||||||||:|||||||||
Db    58  YVASAICLVFMIGEIIGGYLAQSLAIMTDAAHLLTDFASMLISLFSLWVSSRPATKTMNF  117

Qy   135  GWQRAEILGALVSVLSIWVVTGVLVYLAVERLISGDYEIDGGTMLITSGCAVAVNIIMGL  194
          ||||||||||||:||||||||||||||||:|||||||||:|.|||||||||||||||||||
Db   118  GWQRAEILGALLSVLSIWVVTGVLVYLAVQRLISGDYEIKGDTMLITSGCAVAVNIIMGL  177

Qy   195  TLHQSGHGHSHG-----TINQQEENPSVRAAFIHVIGDFMQSMGVLVAAYILYFKPEYKYV  250
          :||||||||||||        :||:|||||||||||:||::||:|||||||:|||||||||
Db   178  ALHQSGHGHSHGHSHEDSSQOQONPSVRAAFIHVVGDLLQSVGVLVAAYIIYFKPEYKYV  237

Qy   251  DPICTFVFSILVLGTTLTILRDVILVLMEGTPKGVDFTAVRDLLLSVEGVEALHSLHWA  310
          |||||:|||||||||||||||||||||||||||||||.||:.:|||||:|||||||||||
Db   238  DPICTFLFSILVLGTTLTILRDVILVLMEGTPKGVDFTTVKNLLLSVDGVEALHSLHWA  297

Qy   311  LTVAQPVLSVHIAIAQNTDAQQAVLKTASSRLQGKFHFHVTIQIEDYSEDMKDCQACQGP  370
          |||||||||||||||||.||.|||||.|..||||||.||.:||||.||||||.||.|.||
Db   298  LTVAQPVLSVHIAIAQNVDADQAVLKVARDRLQGKFNFHHMTIQIESYSEDMKSCQEGCGP  357

Qy   371  SD  372
          ||
Db   358  SE  359
```

RESULT 2

```
ID   ZNT3_MOUSE   STANDARD;   PRT;   388 AA.
AC   P97441; P97511;
DT   30-MAY-2000 (Rel. 39, Created)
DT   30-MAY-2000 (Rel. 39, Last sequence update)
DT   15-JUN-2002 (Rel. 41, Last annotation update)
DE   Zinc transporter 3 (ZnT-3).
GN   SLC30A3 OR ZNT3.
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX   NCBI_TaxID=10090;
RN   [1]
RP   SEQUENCE FROM N.A., AND CHARACTERIZATION.
RX   MEDLINE=97121493; PubMed=8962159;
RA   Palmiter R.D., Cole T.B., Quaife C.J., Findley S.D.;
RT   "ZnT-3, a putative transporter of zinc into synaptic vesicles.";
RL   Proc. Natl. Acad. Sci. U.S.A. 93:14934-14939(1996).
CC   -!- FUNCTION: INVOLVED IN ACCUMULATION OF ZINC IN SYNAPTIC VESICLES.
CC   -!- SUBCELLULAR LOCATION: Integral membrane protein.
CC   -!- LOCALIZED IN THE MEMBRANE OF SYNAPTIC VESICLES (PROBABLE).
CC   -!- TISSUE SPECIFICITY: BRAIN AND TESTIS. IN THE BRAIN, MOST ABUNDANT
CC       IN HIPPOCAMPUS AND CEREBRAL CORTEX. IN THE TESTIS, EXPRESSION IS
CC       RESTRICTED TO GERM CELLS AND IS HIGHEST IN PACHYTENE SPERMATOCYTES
CC       AND ROUND SPERMATIDS.
CC   -!- SIMILARITY: BELONGS TO THE CATION DIFFUSION FACILITATOR FAMILY OF
CC       TRANSPORTERS (CDF; TC 2.A.4). SLC30A SUBFAMILY.
```

```
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   -------------------------------------------------------------------------
DR   EMBL; U76007; AAB39731.1; -.
DR   EMBL; U76009; AAB39733.1; ALT_SEQ.
DR   EMBL; U76008; AAB39733.1; JOINED.
DR   MGD; MGI:1345280; Slc30a3.
DR   InterPro; IPR002524; Cation_efflux.
DR   Pfam; PF01545; Cation_efflux; 1.
DR   TIGRFAMs; TIGR01297; CDF; 1.
KW   Zinc transport; Transport; Transmembrane; Multigene family.
FT   DOMAIN        1     75       CYTOPLASMIC (POTENTIAL).
FT   TRANSMEM     76     96       POTENTIAL.
FT   DOMAIN       97    105       VACUOLAR (POTENTIAL).
FT   TRANSMEM    106    126       POTENTIAL.
FT   DOMAIN      127    145       CYTOPLASMIC (POTENTIAL).
FT   TRANSMEM    146    166       POTENTIAL.
FT   DOMAIN      167    177       VACUOLAR (POTENTIAL).
FT   TRANSMEM    178    198       POTENTIAL.
FT   DOMAIN      199    235       CYTOPLASMIC (POTENTIAL).
FT   TRANSMEM    236    256       POTENTIAL.
FT   DOMAIN      257    263       VACUOLAR (POTENTIAL).
FT   TRANSMEM    264    284       POTENTIAL.
FT   DOMAIN      285    388       CYTOPLASMIC (POTENTIAL).
SQ   SEQUENCE   388 AA;  41824 MW;  3CCDD0A37074EF41 CRC64;

Query Match             52.3%;
Best Local Similarity   52.6%;
Matches   203;  Conservative   52;  Mismatches  102;  Indels   29;  Gaps    6;
                                        Score 1006.5;  DB 1;  Length 388;
                                        Pred. No. 1.6e-67;

Qy    5 EKQHLLDARPAIRSYTG--------SLWQEGAGWIPLPRPGCLDLQAIELAAQSNHHCHAQKG 58
          |: ||| |||        |||:||  |  |     |  |||||||||||| :|:|||| |
Db   11 ETTRLVSARD--RSSAGGGLRLKSLFTEPSE--PLP----EEPKLEGMAFHHCHKDPV 60

Qy   59 PDSHCDPKKGKAQRQLYVASACLLFMIGEVVGGYLAHSLAVMTDAAHLLTDFASMLISL 118
        | : ||||  :||||||| ||| |||| |||||||||||| ||||||||| |:|||:||
Db   61 PQSGLSPERVQARRQLYAACAVCFIFMAGEVVGGYLAHSLAIMTDAAHLLADIGSMLASL 120

Qy  119 FSLWMSSRPATKTMNFGWQRAEILGALVSLVSLSIWVTGVLVLYLAVERLISGDYEIDGGTM 178
        ||||:|:|||| | :|||:|||:|||| |:|||::|||:|:||||| :|| ||:|:|: | |
Db  121 FSLWLSTRPATRTRMTFGWHRSETLGALASVVSLWIVTGILYLYLAFLRLLHSDYHIEAGAM 180

Qy  179 LITSGCAVAVNIIMGLTLHQSGHGHSHGTTNQQ----EE------NPSVRAAFIHVI 225
        |||: ||  |:|:||||||::|   |||   ::|    ||       | :|||||||:
Db  181 LITASIAVCANLLMAFVLHQTGAPHSHGSTGAEYAPLEEGHGYPMSLGNTSVRAAFVHVL 240

Qy  226 GDEMQSMGVLVAAYILYFKPEYKYKVVDPICTFVPSILVLGTTLTLIDRDVILVLMEGTPKGV 285
        || :||:|||| | ||||||:|| ||.|| |:|: ||:||||| ||:||:|||||||| |:|
Db  241 GDLLQSFGVLAASILIYFKPQYKVADPISTFLFSICALGSTAPTLRDVLLVLMEGGAPRSV 300

Qy  286 DFTAVRDLLLSVEGVEALHSLHIWALTVAQPVLSVHIAIAQNTDAQAVLKTASSRLQGKF 345
        :| :|||:||||.||.| |.||:|||| ::.| ||::||:.|:|:||||..|||||:.:|
Db  301 EFEPVRDTLLSVPGVRATHDLHLWALTLTYHVASAHLAIDSTADPEAVLAEASSRLYSRF 360

Qy  346 HFHTVTIQIEDYSEDMKDCQACQGPS 371
        .|.:.||| ::|  :||||: ||.||
Db  361 GFSSCTLQVEQYQPEMAQCLRCQEPS 386
```

RESULT 3

```
ZNT3_HUMAN
```

```
CC   -!- FUNCTION: INVOLVED IN ACCUMULATION OF ZINC IN SYNAPTIC VESICLES
CC       (BY SIMILARITY).
CC   -!- SUBCELLULAR LOCATION: Integral membrane protein.
CC       LOCALIZED IN THE MEMBRANE OF SYNAPTIC VESICLES (PROBABLE).
CC   -!- SIMILARITY: BELONGS TO THE CATION DIFFUSION FACILITATOR FAMILY OF
CC       TRANSPORTERS (CDF; TC 2.A.4). SLC30A SUBFAMILY.
```

```
DR   EMBL; U76010; AAB39732.1; -.
DR   Genew; HGNC:11014; SLC30A3.
DR   MIM; 602878;
DR   InterPro; IPR002524; Cation_efflux.
DR   Pfam; PF01545; Cation_efflux; 1.
DR   TIGRFAMs; TIGR01297; CDF; 1.
KW   Zinc transport; Transport; Transmembrane; Multigene family.
FT   DOMAIN        1     75       CYTOPLASMIC (POTENTIAL).
FT   TRANSMEM     76     96       POTENTIAL.
FT   DOMAIN       97    105       VACUOLAR (POTENTIAL).
FT   TRANSMEM    106    126       POTENTIAL.
FT   DOMAIN      127    145       CYTOPLASMIC (POTENTIAL).
FT   TRANSMEM    146    166       POTENTIAL.
FT   DOMAIN      167    177       VACUOLAR (POTENTIAL).
FT   TRANSMEM    178    198       POTENTIAL.
FT   DOMAIN      199    235       CYTOPLASMIC (POTENTIAL).
FT   TRANSMEM    236    256       POTENTIAL.
FT   DOMAIN      257    264       VACUOLAR (POTENTIAL).
FT   TRANSMEM    265    285       POTENTIAL.
FT   DOMAIN      286    388       CYTOPLASMIC (POTENTIAL).
SQ   SEQUENCE   388 AA;  41944 MW;  74BB24F17CF4740 CRC64;

Query Match             50.6%;
Best Local Similarity   53.7%;
Matches   189;  Conservative   50;  Mismatches   94;  Indels   19;  Gaps    4;
                                        Score 972.5;  DB 1;  Length 388;
                                        Pred. No. 5.2e-65;

Qy   32 PLPRPGCLDLQAIELAAQSNHHCHAQKGPDSHCDPKKGKAQRQLYVASACLLFMIGEVVG 91
        ||| |   |   :   |||:||:||| |||:|||:: ::||||||||| ||| |||| |||
Db   40 PLPE---ESKPVEMPF--HHCHRDPLPPGCLTPERLHARRQLYAACAVCFVFMAGEVVG 93

Qy   92 GYLAHSLAVMTDAAHLLTDFASMLISLFSLWMSSRPATKTMNFGWQRAEILGALVSLVSLSI 151
        |||||||| ||||||||| ||| |||| ||||:||||||:|:|||:||:|||| |:|||::
Db   94 GYLAHSLAIMTDAAHLLADVGSMMGSLFSLWLSTRPATRTMTFGWHRSETLGALASVVSL 153

Qy  152 WVTGVLVLYLAVERLISGDYEIDGGTMLITSGCAVAVNIIMGLTLHQSGHGHSH 211
        |:||:| ||||| :|| ||:|:| | |||: ||  |:|:||||||::|  |||
Db  154 WMVTGILYLYLAFVRLLHSDYHIEGGAMLLTASIAVCANLLMAFVLHQAGPPHSHGSRGAE 213

Qy  212 ----EENP-------SVRAAFIHVIGDFMQSMGVLVAAYILYFKPEYKYKVVDPICTFVP 258
            ||         :|||||:||:||  ||:|||| | |||||:|| ||.|| |:|:
Db  214 YAPLEEGPEQPLPLGNTSVRAAFVHVLGDLLQSFGVLAASILIYFKPQYKVAADPISTFLF 273

Qy  259 SILVLGTTLTLIDRDVILVLMEGTPKGVDFTAVRDLLLSVEGVEALHSLHIWALTVAQPVL 318
        ||:||||| ||:||:||||||||||.| :|| |||:||||.||.| |.||:||||:::.|
Db  274 SICALGSTAPTLRDVLRILMEGTPRNVGFEPVRDTLLSVPGVRATHELHLWALTLTYHVA 333

Qy  319 SVHIAIAQNTDAQAVLKTASSRLQGKFHFHTVTIQIEDYSEDMKDCQACQGP 370
        | |:||:.|:|:||||..|||||:.:|.|.:.||| ::|  :||||: ||
Db  334 SAHLAIDSTADPEAVLAEASSRLYSRFGFSSCTLQVEQYQPEMAQCLRCQEP 385
```

RESULT 4

ID	ZNT4_HUMAN	STANDARD;	PRT; 429 AA.
AC	O14863;		
DT	30-MAY-2000 (Rel. 39, Created)		
DT	30-MAY-2000 (Rel. 39, Last sequence update)		
DT	15-JUN-2002 (Rel. 41, Last annotation update)		
DE	Zinc transporter 4 (ZnT-4).		
GN	SLC30A4 OR ZNT4.		
OS	Homo sapiens (Human).		
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
OC	Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.		
OX	NCBI_TaxID=9606;		
RN	[1]		
RP	SEQUENCE FROM N.A.		
RC	TISSUE=Fetal brain;		
RX	MEDLINE=98016412; PubMed=9354792;		
RA	Huang L., Gitschier J.;		
RT	"A novel gene involved in zinc transport is deficient in the lethal		
RT	milk mouse.";		
RL	Nat. Genet. 17:292-297(1997).		
CC	-!- FUNCTION: PROBABLY INVOLVED IN ZINC TRANSPORT OUT OF THE		
CC	CYTOPLASM, MAY BE BY SEQUESTRATION INTO AN INTRACELLULAR		
CC	COMPARTMENT.		
CC	-!- SUBCELLULAR LOCATION: Integral membrane protein.		
CC	LOCALIZED IN ENDOSOMAL VESICLE MEMBRANE (PROBABLE).		
CC	-!- DOMAIN: CONTAINS A HISTIDINE-RICH REGION WHICH IS A LIGAND FOR		
CC	ZINC AND AN ASPARTIC ACID-RICH REGION WHICH IS A POTENTIAL LIGAND		
CC	FOR ZINC (BY SIMILARITY)		
CC	-!- SIMILARITY: BELONGS TO THE CATION DIFFUSION FACILITATOR FAMILY OF		
CC	TRANSPORTERS (CDF; TC 2.A.4). SLC30A SUBFAMILY.		
CC	--		
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration		
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation -		
CC	the European Bioinformatics Institute. There are no restrictions on its		
CC	use by non-profit institutions as long as its content is in no way		
CC	modified and this statement is not removed. Usage by and for commercial		
CC	entities requires a license agreement (see http://www.isb-sib.ch/announce/		
CC	or send an email to license@isb-sib.ch).		
CC	--		
DR	EMBL; AF025409; AAB82561.1; -.		
DR	Genew; HGNC:11015; SLC30A4.		
DR	MIM; 602095; -.		
DR	InterPro; IPR002524; Cation_efflux.		
DR	Pfam; PF01545; Cation_efflux; 1.		
DR	TIGRFAMs; TIGR01297; CDF; 1.		
KW	Zinc transport; Transport; Transmembrane; Multigene family.		
FT	DOMAIN 1 113 CYTOPLASMIC (POTENTIAL).		
FT	TRANSMEM 114 134 POTENTIAL.		
FT	DOMAIN 135 143 VACUOLAR (POTENTIAL).		
FT	TRANSMEM 144 164 POTENTIAL.		
FT	DOMAIN 165 178 CYTOPLASMIC (POTENTIAL).		
FT	TRANSMEM 179 199 POTENTIAL.		
FT	DOMAIN 200 216 VACUOLAR (POTENTIAL).		
FT	TRANSMEM 217 237 POTENTIAL.		
FT	DOMAIN 238 274 CYTOPLASMIC (POTENTIAL).		
FT	TRANSMEM 275 295 POTENTIAL.		
FT	DOMAIN 296 310 VACUOLAR (POTENTIAL).		
FT	TRANSMEM 311 331 POTENTIAL.		
FT	DOMAIN 332 429 CYTOPLASMIC (POTENTIAL).		
FT	DOMAIN 17 96 HIS-RICH.		
FT	DOMAIN 241 247 ASP-RICH (ACIDIC).		
SQ	SEQUENCE 429 AA; 47496 MW; 53892DFF8768739 CRC64;		

```
Qy      2  EAKEKQHLLDARPAIRSYTGSLWOEGAGWIPLPRPGLDLQAIELAAQSNHHCHAQKGPDS 61
           |||||||||||||||| ::||:|||  ||:|| |||| ||:|: | ||:| |:||:||
Db     56  EAPERP-VNGAHPTLQADDDSLLDQD----LPLTNSQLSLK----VDSCDNCSKQRE--- 103
```

RESULT 5

ID	ZNT4_MOUSE	STANDARD;	PRT; 430 AA.
AC	O35149; O35154;		
DT	30-MAY-2000 (Rel. 39, Created)		
DT	30-MAY-2000 (Rel. 39, Last sequence update)		
DT	15-JUN-2002 (Rel. 41, Last annotation update)		
DE	Zinc transporter 4 (ZnT-4) (Lethal milk protein).		
GN	SLC30A4 OR ZNT4 OR LM.		
OS	Mus musculus (Mouse).		
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
OC	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.		
OX	NCBI_TaxID=10090;		
RN	[1]		
RP	SEQUENCE FROM N.A.		
RC	STRAIN=C57BL/6, and B6/CBAF1; TISSUE=Brain;		
RX	MEDLINE=98016412; PubMed=9354792;		
RA	Huang L., Gitschier J.;		
RT	"A novel gene involved in zinc transport is deficient in the lethal		
RT	milk mouse.";		
RL	Nat. Genet. 17:292-297(1997).		
CC	-!- FUNCTION: PROBABLY INVOLVED IN ZINC TRANSPORT OUT OF THE		
CC	CYTOPLASM, MAY BE BY SEQUESTRATION INTO AN INTRACELLULAR		
CC	COMPARTMENT.		
CC	-!- SUBCELLULAR LOCATION: Integral membrane protein.		
CC	LOCALIZED IN ENDOSOMAL VESICLE MEMBRANE (PROBABLE).		
CC	-!- TISSUE SPECIFICITY: WIDELY EXPRESSED. HIGHLY EXPRESSED IN THE		
CC	BRAIN AND IN MAMMARY EPITHELIAL CELL LINES.		
CC	-!- DOMAIN: CONTAINS A HISTIDINE-RICH REGION WHICH IS A LIGAND FOR		
CC	ZINC AND AN ASPARTIC ACID-RICH REGION WHICH IS A POTENTIAL LIGAND		
CC	FOR ZINC (BY SIMILARITY).		
CC	-!- DISEASE: LETHAL MILK (LM) MICE ARE DEFECTIVE IN ZINC TRANSPORT		
CC	INTO BREAST MILK, DUE TO A PREMATURE TRANSLATION TERMINATION CODON		
CC	AT POSITION 297. ONLY HOMOZYGOUS MUTANT ADULTS DEVELOP DERMATITIS,		
CC	SKIN LESIONS, AND HAIR LOSS DUE TO A SYSTEMIC ZINC DEFICIENCY.		
CC	HOWEVER NEONATAL MICE (OF ANY GENOTYPE) SUCKLED ON HOMOZYGOUS		
CC	MUTANT FEMALE ALSO DEVELOP SYMPTOMS CHARACTERISTIC OF NUTRITIONAL		
CC	ZINC DEFICIENCY, INCLUDING DERMATITIS, ALOPECIA AND STUNTED		
CC	GROWTH.		
CC	-!- SIMILARITY: BELONGS TO THE CATION DIFFUSION FACILITATOR FAMILY OF		
CC	TRANSPORTERS (CDF; TC 2.A.4). SLC30A SUBFAMILY.		
CC	--		
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration		
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation -		
CC	the European Bioinformatics Institute. There are no restrictions on its		
CC	use by non-profit institutions as long as its content is in no way		
CC	modified and this statement is not removed. Usage by and for commercial		
CC	entities requires a license agreement (see http://www.isb-sib.ch/announce/		

```
Query Match    37.6%;  Score 722.5;  DB 1;  Length 429;
Best Local Similarity   39.0%;  Pred. No. 1.9e-46;
Matches 151; Conservative 88; Mismatches 117; Indels 31; Gaps 7;

Qy     62  HCDPKKGKAQRQLYVASAICLLFMIGEVVGGYLAHSLAVMTDAAHLLTDFASMLISLFSL 121
           |:|: :|.::| || :||||||| ||:||.:| :|:|| :::||:.|| ::|||: |:|
Db    104  --ILKQRVKARLTIAAVLYLLFMIGELVGGYIANSLAIMTDALHMLTDLSAIILTLLAL 161

Qy    122  WMSSRPATKTMNFGWQRAEILGALVSVLSIWVVTGVLVYLAVEBLISGDYEIDGGTMLIT 181
           | ||| |:: :|  |:|  |: | |||:|| :||:||||:|| |   |||:||  ||||
Db    162  WLSSKSPFIKRFTFGFHRLEVLSAMISVLLVYILMGFLLYEAVQRTIHMNYEINGDIMLIT 221

Qy    182  SGCAVANIIMGLTLHQSGHSHGHTT--------------NQQEENPSVRRAFIHVIGD 227
           :||:|| ||:||: :| |:|:|| |            ::||||||:|||:|||:|| ||
Db    222  AAVGVAVNVIMGFLLNQSGHRSHSHSHSLPSNSPTRGSGCERNHGQDSLAVRRAFVHALGD 281

Qy    228  FMQSMGVLIVAAYILYFKPEYKIVDPICTFVFSIILVLGTTLTLILRDVILVLMEGTPKGVDF 287
           | |:|||::|||:||  |:|| | ||:| || |||:|||:|:::||||| || :||| |||
Db    282  LVQSVGVLIAAYIIRFKPEYKIADPICIYVFSLLVAFTTFRIIWDTVVIILEGVPSHLNV 341

Qy    288  TAVRDLLLSVEGVEALHSLHIWALTVAQPVLSVHIAIAQNTDA--QAVLKTASSRLQGKF 345
           ::||:|: |:||:|:|| ||| | ||| |::| |||:|:::  :|| || |||:|| ||
Db    342  DYIKEALMKIEDVYSVEDLNIWSLTSGKSTAIVHIQLLIPGSSSKWEEVQSKANHLLLNTF 401

Qy    346  HFHTVTLQIEDYSEDM-KDCQACQGPS 371
           :|||::| :|| | |  |:|:|: | |:
Db    402  GMYRCTIQLQSYRQEVDRTCANCQSSS 428
```

CC or send an email to license@isb-sib.ch).
CC --
DR EMBL; AF004099; AAB82413.1; -.
DR EMBL; AF003747; AAB82593.1; -.
DR EMBL; AF004098; AAB82411.1; -.
DR EMBL; AF004097; AAB82412.1; -.
DR EMBL; AF004100; AAB82414.1; -.
MGD; MGI:1345282; Slc30a4.
DR InterPro; IPR002524; Cation_efflux.
DR MGD; MGI:1345282; Slc30a4.
Pfam; PF01545; Cation_efflux; 1.
DR TIGRFAMs; TIGR01297; CDF; 1.
KW Zinc transport; Transport; Transmembrane; Multigene family.
FT DOMAIN   1   113   CYTOPLASMIC (POTENTIAL).
FT TRANSMEM   114   134   POTENTIAL.
FT DOMAIN   135   143   VACUOLAR (POTENTIAL).
FT TRANSMEM   144   164   POTENTIAL.
FT DOMAIN   165   178   CYTOPLASMIC (POTENTIAL).
FT TRANSMEM   179   199   POTENTIAL.
FT DOMAIN   200   216   VACUOLAR (POTENTIAL).
FT TRANSMEM   217   237   POTENTIAL.
FT DOMAIN   238   275   CYTOPLASMIC (POTENTIAL).
FT TRANSMEM   276   296   POTENTIAL.
FT DOMAIN   297   311   VACUOLAR (POTENTIAL).
FT TRANSMEM   312   332   POTENTIAL.
FT DOMAIN   333   430   CYTOPLASMIC (POTENTIAL).
FT DOMAIN   17   96   ASP-RICH (ACIDIC).
FT DOMAIN   241   249   HIS-RICH.
FT CONFLICT   426   426   Q -> H (IN REF. 1).
SQ SEQUENCE   430 AA; 47791 MW; E231IFA625D15C6 CRC64;

Query Match   37.2%; Score 716; DB 1; Length 430;
Best Local Similarity   37.7%; Conservative   89; Mismatches 114; Indels 42; Gaps   7;
Matches 148;

Qy   2 EAKEKQHLLDARPAIRSYTGSLWQEGAGWIPLPRPGLDLQAIELAAQSNHHCHAQKGPDS 61
         |||| : |:|::|||:||||||||||||:|||||:||||||:||||:|||||||
Db   56 EAPERP-VNGAHPALQADDDSLLDQD---LPLTNSQLSL----------KMDPCD 96

Qy   62 HCDP-----KKGKAQRQLYVASAICLLFMIGEVGGYIAHSLAVMTDAAHLLTDFASMLI 116
         :| |:||||||||||||||||||||||||||||||||||||||||||||||
Db   97 NCSKRRELLKQRKVKTRLTIAAVLYLLFMIGELVGGYMANSLAIMTDALHMLTDLSAIL 156

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxID=10116;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Wistar; TISSUE=Brain;
RX MEDLINE=20068535; PubMed=10600821;
RA Murgia C., Vespignani I., Cerase J., Nobili F., Perozzi G.;
RT "Cloning, expression, and vesicular localization of zinc transporter
RT Dri 27/ZnT-4 in intestinal tissue and cells.";
RL Am. J. Physiol. 277:G1231-G1239(1999).
-!- FUNCTION: PROBABLY INVOLVED IN ZINC TRANSPORT OUT OF THE
CC   CYTOPLASM, MAY BE BY SEQUESTRATION INTO AN INTRACELLULAR
CC   COMPARTMENT.
-!- SUBUNIT: MEDIATES HETERODIMERIC INTERACTIONS WITH AT LEAST ONE
CC   SPECIFIC PARTNER.
-!- SUBCELLULAR LOCATION: Integral membrane protein.
CC   LOCALIZED IN ENDOSOMAL VESICLE MEMBRANE (PROBABLE), IN THE
CC   POLARIZED ENTEROCYTES, IT IS MAINLY LOCALIZED IN THE BASAL
CC   CYTOPLASMIC REGION.
-!- TISSUE SPECIFICITY: WIDELY EXPRESSED. HIGHLY EXPRESSED IN BRAIN
CC   AND TESTES. ALSO EXPRESSED IN SMALL INTESTINE, MEDULLA, LUNG,
CC   KIDNEY, STOMACH, AND COLON. EXPRESSED AT LOWER LEVEL IN OTHER
CC   TISSUES.
-!- DEVELOPMENTAL STAGE: DEVELOPMENTALLY REGULATED IN THE INTESTINE.
-!- INDUCTION: NO CHANGE IN RESPONSE TO ZINC DEPRIVATION.
-!- DOMAIN: CONTAINS A HISTIDINE-RICH REGION WHICH IS A LIGAND FOR
CC   ZINC AND AN ASPARTIC ACID-RICH REGION WHICH IS A POTENTIAL LIGAND
CC   FOR ZINC.
-!- SIMILARITY: BELONGS TO THE CATION DIFFUSION FACILITATOR FAMILY OF
CC   TRANSPORTERS (CDF, TC 2.A.4). SLC30A SUBFAMILY.
CC --
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC --
DR EMBL; Y16774; CAA76372.1; -.
DR InterPro; IPR002524; Cation_efflux.
Pfam; PF01545; Cation_efflux; 1.
DR TIGRFAMs; TIGR01297; CDF; 1.
KW Zinc transport; Transport; Transmembrane; Multigene family.
FT DOMAIN   1   113   CYTOPLASMIC (POTENTIAL).
FT TRANSMEM   114   134   POTENTIAL.
FT DOMAIN   135   143   VACUOLAR (POTENTIAL).
FT TRANSMEM   144   164   POTENTIAL.
FT DOMAIN   165   178   CYTOPLASMIC (POTENTIAL).
FT TRANSMEM   179   199   POTENTIAL.
FT DOMAIN   200   216   VACUOLAR (POTENTIAL).
FT TRANSMEM   217   237   POTENTIAL.
FT DOMAIN   238   275   CYTOPLASMIC (POTENTIAL).
FT TRANSMEM   276   296   POTENTIAL.
FT DOMAIN   297   311   VACUOLAR (POTENTIAL).
FT TRANSMEM   312   332   POTENTIAL.
FT DOMAIN   333   430   CYTOPLASMIC (POTENTIAL).
FT DOMAIN   17   96   ASP-RICH (ACIDIC).
FT DOMAIN   241   249   HIS-RICH.
SQ SEQUENCE   430 AA; 47702 MW; F34CED3FA4F05FB CRC64;

Query Match   37.0%; Score 711; DB 1; Length 430;
Best Local Similarity   38.2%; Conservative   85; Mismatches 116; Indels 42; Gaps   8;
Matches 150;

Qy   2 EAKEKQHLLDARPAIRSYTGSLWQEGAGWIPLPRPGLDLQAIELAAQSNHHCHAQKGPDS 61
         |||| : |:|::|||:||||||||||||:|||||:||||||:||||:|||||||
Db   56 EAPERP-VNGAHPALQADDDSLLDQE---LPLTNSQLSL----------KMDPCD 96

Qy   62 HCDP-----KKGKAQRQLYVASAICLLFMIGEVGGYIAHSLAVMTDAAHLLTDFASMLI 116
         :| |:||||||||||||||||||||||||||||||||||||||||||||||
Db   97 NCSKRRELLKQRKVKTRLTIAAVLYLLFMIGELVGGYMANSLAIMTDALHMLTDLSAIL 156

RESULT 6
ID ZNT4_RAT   STANDARD;   PRT;   430 AA.
AC O55174;
DT 30-MAY-2000 (Rel. 39, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Zinc transporter 4 (ZnT-4) (Dri 27 protein).
GN SLC30A4 OR ZNT4.
OS Rattus norvegicus (Rat).

Qy 340 RLQGKFHFHTVTIQIEDYSED-MKDCQACQGPS 371
          | : : ||:|||:| |:: ||:|:::||
Db 397 LLINTFGMYKCTIQLQSYRQEVIRTCANCQSSS 429

Qy 337 PSHLNVDIKESLMKIEDVYSVEDLNIWSLTSGCKSTAIVHMQLIPGSSSKWEEVQSRAKH 396
          | :||| |::::: : ||| |
Db 337 VHALGDLVQSVGVLIAAVIIRRKPEYKIADPICTYIFSLLVAFTTFRIIWDIVVILEGV 336

Qy 282 PKGVDFTAVRDLLLSVEGVEALHSLHIWALTVAQPVLSVHIAIAQNTDA--QAVLKTASS 339
          |: |: |:||||| ||: |::|| ||::|||||| |||
Db 277 VHALGDLVQSVGVLIAAVIIRRKPEYKIADPICTYIFSLLVAFTTFRIIWDIVVILEGV 336

Qy 222 IHVIGDFMQSMGVLVAAYLYFRPEYKYVDPICTFVFSILVLGTTLTLTLLRDVLVLMEGT 281
         :||: |||: ||||:||| |||:: |||||::|:::||| |: ||:||:|
Db 281 ...

Qy 217 VMLITAAVGVAVNVIMGFLLNQSGHHSHSHSLPSNSPSMVSSGHHHGQDSLAVRRAF 276
         ||:|||||||||:|||| ||| |: |:|||:|
Db 276 ...

Qy 177 TMLITSGCAVAVNIIMGLTLHQSGHGHSHGTT-----------NQQEENPSVRAAF 221
         |||:|: |:||||:|||| |||||:|:|:|
Db 221 ...

Qy 157 TLIALWLSSKSPIRRFTPGFHRLEVLSAMISVMLVYVLMGFLLYEAVQRTIHMNYEINGD 216
         ||:|||:| :: |::| : :|:|:| |:|:|:|| :| |||||||::||::|:|:|
Db 216 ...

Qy 117 SLFSLWMSSRPATKTMNFGWQRAEILGALVSVLSIVVTGVLVYLYLAVERLISGDYEIDGG 176
         ::|:| |||| |::|| | :: ||||| |:||:|:| | |: |: |||:|||||||||
Db 176 ...

Qy   97 NCSKRRELLKQRKVKTRLTIAAVLYLLFMIGELVGGYMANSLAIMTDALHMLTDLSAIL 156
         :| |:||||||||||||||||||||||||||||||||||||||||||||||||:||||
Db 156 ...

Qy   62 HCDP-----KKGKAQRQLYVASAICLLFMIGEVGGYIAHSLAVMTDAAHLLTDFASMLI 116
Db   56 EAPERP-VNGAHPALQADDDSLLDQD---LPLTNSQLSL----------KMDPCD 96

```
RESULT 7
ZITB_SALTY
ID   ZITB_SALTY    STANDARD;    PRT;    312 AA.
AC   Q8ZQT3;
DT   15-JUN-2002 (Rel. 41, Created)
DT   15-JUN-2002 (Rel. 41, Last sequence update)
DT   15-JUN-2002 (Rel. 41, Last annotation update)
DE   Zinc transporter zitB.
GN   ZITB OR STM0758.
OS   Salmonella typhimurium.
OC   Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC   Salmonella.
OX   NCBI_TaxID=602;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=LT2 / SGSC1412 / ATCC 700720;
RX   MEDLINE=21534946; PubMed=11677609;
RA   McClelland M., Sanderson K.E., Spieth J., Clifton S.W., Latreille P.,
RA   Courtney L., Porwollik S., Ali J., Dante M., Du F., Hou S., Layman D.,
RA   Leonard S., Nguyen C., Scott K., Holmes A., Grewal N., Mulvaney E.,
RA   Ryan E., Sun H., Florea L., Miller W., Stoneking T., Nhan M.,
RA   Waterston R., Wilson R.K.;
RT   "Complete genome sequence of Salmonella enterica serovar Typhimurium
RT   LT2.";
RL   Nature 413:852-856(2001).
CC   -!- FUNCTION: Involved in zinc efflux across the cytoplasmic membrane,
CC       thus reducing zinc accumulation in the cytoplasm and rendering
CC       bacteria more resistant to zinc. It may contribute to zinc
CC       homeostasis at low concentrations of zinc (By similarity).
CC   -!- SUBCELLULAR LOCATION: Integral membrane protein (Potential).
CC   -!- SIMILARITY: BELONGS TO THE CATION DIFFUSION FACILITATOR FAMILY OF
CC       TRANSPORTERS (CDF; TC 2.A.4). SLC30A SUBFAMILY.
CC   ----------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ----------------------------------------------------------------------
DR   EMBL; AE008731; AAL19697.1; -.
DR   StyGene; SG?????; zitB.
DR   TIGRFAMs; TIGR01297; CDF; 1.
KW   Transport; Zinc transport; Transmembrane; Complete proteome.
FT   TRANSMEM     21     41       POTENTIAL.
FT   TRANSMEM     48     68       POTENTIAL.
FT   TRANSMEM     90    110       POTENTIAL.
FT   TRANSMEM    123    143       POTENTIAL.
FT   TRANSMEM    164    184       POTENTIAL.
FT   TRANSMEM    240    312       POTENTIAL.
FT   DOMAIN                       HIS-RICH.


Qy  117  SLFSLWMSSRPATKTMNFGWQRAEILGALVSVLSIWVTGVLVLVYLAVERLISGDYEIGG  176
         :|:||:|| |:|: |||||||:|:. : ||||||:|::||:.::|:||| |:|||:
Db  157  TLLALWLSSKSPTRRTFGFPHRLEVLSAMISVMLVVYLMGFLLYEAMQRTIHMNYEINGD  216

Qy  177  TMLITSGCAVANVIIMGLTLHQSGHGHSHGTTNQQEEN-PS----------VRAAF      221
         .|||| . .||||:||| |:||||||.||.||.     ||          ||||
Db  217  VMLITRAAVGAVANVIMGFLLNQSGHGHHHSHAHSHSLPSNSPSMVSSGHSHGQDSLAVRAAF  276

Qy  222  IHVIGDFMQSMGVLVAAYILYFKPEYKVDPICTFVFSILVLGTTLIFTLLRDVILVLMEGT  281
         :|.:||::||.||::||||: ||||||:.|||| :||:||  |::: ::|| ::::| ||
Db  277  VHALGDLVQSVGVLIAAYIIRFKPEYKIADPICTYIFSLLVAFTLTRIIWDTVVIILEGV   336

Qy  282  PKGVDFTAVRDLLLSVEGVEALHSLHTWALTVAQPVLSVHIAIAQNTDA--QAVLKTASS   339
         |..::.:  :.:|:..|.||::.:  ||.|| .:..|:|||: :         |:|:
Db  337  PSHLNVDYIKESLMKIEDVYSVEDLNIWSLTSGKATAIVHMQLIPGSSSKWEEVQSKAKH  396

Qy  340  RLQGKFHFHVTIQIEDY-SEDMKDCQACQGPS                              371
         .|  .| . | .||..| .||..:|  | .:|
Db  397  LLLNTFGMYKCIVQLQSYRQEATRTCANCQSSS                            429
```

```
SQ   SEQUENCE   312 AA;   34461 MW;   2521FF85B2B7B3A6 CRC64;

Query Match              22.6%;   Score 434;   DB 1;   Length 312;
Best Local Similarity    33.1%;   Pred. No. 3e-25;
Matches  107;   Conservative  67;   Mismatches 121;   Indels  28;   Gaps   9;

Qy   49  SNHHCHAQRGPDSHCDPKGKGKAQRQLYVASAICLLFMIGEVVGGYLAHSLAVMTDAAHLL  108
         ::.|||: .|||.|      |.  ::|  :||  |||::||||||:||:||:|:|||||:|
Db    2  AHSHSHA-----DSHL-PKDNNARRLLF-AFIVIAGFMLLEVVGGILSGSLALLADAGHML   55

Qy  109  TDFASMLISLFSLWMSSRPATKTMNFGWQRAEILGALVSVLSIWVTGVLVLVYLAVERLIS  168
         || |:|| :||:||.| ::.|.. ||||:|| ||.|:. |::|:| |::||:|::.||::
Db   56  TDAAALLFALLAVQQFSRRPPTVRHTFGWLRLTTLAAFVNAIALVVITLLIVWEAIERFYT  115

Qy  169  GDYEIDGGTMLITSGCAVANVIIMGLTLHQSGHGHSHGTTNQQEENPSVRAAFIHVIGDF   228
         |  .:.|.:.:::.... ||:::. |||:.    .   |.:.     ||||| :||:||:
Db  116  -PRPVAGNLMMVIAVAGLLANLFAFWILHR----------GSDEKNLNVRAAALHVMGDL   164

Qy  229  MQSMGVLVAAYILYFKPEYKVDPICTFVFSILVLGTTLIFTLLRDVILVLMEGTPKGVDFT  288
         :||.| :||| :: | ||.. | | |:.||:|.| : |: |||:.:.|:. . ||..|:.
Db  165  LGSVGAIVAALIIIW-TGWTPADPILSILVSVVLVRSAWRLLKDSVNELLEGAPVSLDIN   223

Qy  289  AV-RDLLLSVEGVEALHSLHIWALTVAQPVLSVHIAIAQNTDAQAVLKTASSRLQGKFHF  347
         |: :.:|::.: :: .|::| :::.   |.:. |..|   | .:||::. |:.:.: :|:
Db  224  ALQRHLSREIPEYRNVHVHVHVM-MVGEKPVMTLHAQVIPPHDHDALLERIQDFLMHEYHI  282

Qy  348  HVTIQIEDYSEDMKDCQACQGP                                        370
         ..|:|::       .:|  |||
Db  283  AHATIQME---------YQVCHGP                                      297
```

```
RESULT 8
ZITB_SALTI
ID   ZITB_SALTI    STANDARD;    PRT;    312 AA.
AC   Q8Z8B6;
DT   15-JUN-2002 (Rel. 41, Created)
DT   15-JUN-2002 (Rel. 41, Last sequence update)
DT   15-JUN-2002 (Rel. 41, Last annotation update)
DE   Zinc transporter zitB.
GN   ZITB OR STY0799.
OS   Salmonella typhi.
OC   Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC   Salmonella.
OX   NCBI_TaxID=601;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=CT18;
RX   MEDLINE=21534947; PubMed=11677608;
RA   Parkhill J., Dougan G., James K.D., Thomson N.R., Pickard D., Wain J.,
RA   Churcher C., Mungall K.L., Bentley S.D., Holden M.T.G., Sebaihia M.,
RA   Baker S., Basham D., Brooks K., Chillingworth T., Connerton P.,
RA   Cronin A., Davis P., Davies R.M., Dowd L., White N., Farrar J.,
RA   Feltwell T., Hamlin N., Haque A., Hien T.T., Holroyd S., Jagels K.,
RA   Krogh A., Larsen T.S., Leather S., Moule S., O'Gaora P., Parry C.,
RA   Quail M., Rutherford K., Simmonds M., Skelton J., Stevens K.,
RA   Whitehead S., Barrell B.G.;
RT   "Complete genome sequence of a multiple drug resistant Salmonella
RT   enterica serovar Typhi CT18.";
RL   Nature 413:848-852(2001).
CC   -!- FUNCTION: Involved in zinc efflux across the cytoplasmic membrane,
CC       thus reducing zinc accumulation in the cytoplasm and rendering
CC       bacteria more resistant to zinc. It may contribute to zinc
CC       homeostasis at low concentrations of zinc (By similarity).
CC   -!- SUBCELLULAR LOCATION: Integral membrane protein (Potential).
CC   -!- SIMILARITY: BELONGS TO THE CATION DIFFUSION FACILITATOR FAMILY OF
CC       TRANSPORTERS (CDF; TC 2.A.4). SLC30A SUBFAMILY.
CC   ----------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
```

```
CC   or send an email to license@isb-sib.ch).
CC   -----
DR   EMBL; AL627268; CAD05215.1; -.
DR   TIGRFAMS; TIGR01297; CDF; 1.
KW   Transport; Zinc transport; Transmembrane; Complete proteome.
FT   TRANSMEM     21     41       POTENTIAL.
FT   TRANSMEM     48     68       POTENTIAL.
FT   TRANSMEM     90    110       POTENTIAL.
FT   TRANSMEM    123    143       POTENTIAL.
FT   TRANSMEM    164    184       POTENTIAL.
FT   DOMAIN      240    312       HIS-RICH.
SQ   SEQUENCE    312 AA;  34531 MW;  EA29F94B3A18736 CRC64;

Query Match             22.4%;  Score 431;  DB 1;  Length 312;
Best Local Similarity   33.3%;
Matches  108;  Conservative   66;  Mismatches  130;  Indels   20;  Gaps    8;

Qy      49 SNHHCHAQKGPDSHCDPKKGKAQORLYVASAICLLFMIGEVVGGYLAHSLAVMIDAAHLL 108
           ::  :  ::.   .     ::::     :|||||||:.::|||||:|||::::|:::||:|
Db       2 AHSHSHA----DSHL-PKDNNARRLLF-AFIVTAGFMLLEVVGGIISGGSIALLADAGHML 55

Qy     109 TDFASMLISLFSLWMSSRPATKTMNFGWQRAEILGALYSVLSISWVVTGVLVYLAVERLIS 168
           ||..::|  :  || |||||.   | ||.:|   |  ::.: :  :||.||::|.||:|. :
Db      56 TDAALLFALLVVQFSRRPPTVRHTFGWLRLTTLAAFVNAIALVVITLLIVWEAIERFYT 115

Qy     169 GDYEIDGGTMLITSGCAVAVNIIMGLTLHQSGHGHSHGTTNQOEENPSVRAAFIHVIGDF 228
           .   |   ::.|:: . || ||:  |::|:   |         ||::||||:||:|||:
Db     116 -PRVAGNLMMVIAVAGLLANLFAFWILHR------GSDEKNLNVRAALHVMGDL 164

Qy     229 MQSMGVLVAAYILYFKPEYKYVDPICTFVFSILVLGTTLFTLRDVILVLMEGTPKGVDFT 288
           : |.|.:|||::. .   .  .|||:.::|.:||||:|:::|:||.::.:|||:|.:.:
Db     165 LGSVGAIVAALIIW-TGWTPADPILSIVSVLVLRSAWRLLKDSVNELLEGAPVSLDIN 223

Qy     289 AV-RDLLLSVEGVEALHSLHIWALTVAQPVLSVHIAIAQNTDAQAVLKTASSRLQGKFHF 347
           |:  .:| ||  |.:||:||:|:  :.|:.: .|.: |::.|   |::.::. : | .|
Db     224 ALQRHLSREIPEVRNVHVHVHVW-MVGEKRVMTLHAQVIPPHDHDALLERIQDFLMHEYHI 282

Qy     348 HTVTIQIEDYSEDMKDCQACQGPS 371
           | .:||:| ||.    |.:|.||:
Db     283 AHAFIQMEYQOMCHGPDCHLNQTPS 306


RESULT 9
ZITB_ECOLI

ID   ZITB_ECOLI     STANDARD;      PRT;   313 AA.
AC   P75757;
DT   01-NOV-1997 (Rel. 35, Created)
DT   01-NOV-1997 (Rel. 35, Last sequence update)
DT   15-JUN-2002 (Rel. 41, Last annotation update)
DE   Zinc transporter ZitB.
GN   ZITB OR B0752.
OS   Escherichia coli.
OC   Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC   Escherichia.
OX   NCBI_TaxID=562;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=K12 / MG1655;
RX   MEDLINE=97426617; PubMed=9278503;
RA   Blattner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,
RA   Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F.,
RA   Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,
RA   Mau B., Shao Y.;
RT   "The complete genome sequence of Escherichia coli K-12.";
RL   Science 277:1453-1474(1997).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=K12;
RX   MEDLINE=97061202; PubMed=8905232;
RA   Oshima T., Aiba H., Baba T., Fujita K., Hayashi K., Honjo A.,
RA   Ikemoto K., Inada T., Itoh T., Kajihara M., Kanai K., Kashimoto K.,
RA   Kimura S., Kitagawa M., Makino K., Masuda S., Miki T., Mizobuchi K.,
RA   Mori H., Motomura K., Nakamura Y., Nashimoto H., Nishio Y., Saito N.,
RA   Sampei G., Seki Y., Tagami H., Takemoto K., Wada C., Yamamoto Y.,
RA   Yano M., Horiuchi T.;
RT   "A 718-kb DNA sequence of the Escherichia coli K-12 genome
RT   corresponding to the 12.7-28.0 min region on the linkage map.";
RL   DNA Res. 3:137-155(1996).
RN   [3]
RP   CHARACTERIZATION.
RC   STRAIN=K12 / W3110;
RX   MEDLINE=21136524; PubMed=11443104;
RA   Grass G., Fan B., Rosen B.P., Franke S., Nies D.H., Rensing C.;
RT   "ZitB (YbgR), a member of the cation diffusion facilitator family, is
RT   an additional zinc transporter in Escherichia coli.";
RL   J. Bacteriol. 183:4664-4667(2001).
CC   -!- FUNCTION: Involved in zinc efflux across the cytoplasmic membrane,
CC       thus reducing zinc accumulation in the cytoplasm and rendering
CC       bacteria more resistant to zinc. It may contribute to zinc
CC       homeostasis at low concentrations of zinc, whereas zntA is
CC       required for growth at more toxic zinc concentrations.
CC   -!- SUBCELLULAR LOCATION: Integral membrane protein (Potential).
CC   -!- INDUCTION: By zinc.
CC   -!- MISCELLANEOUS: Appears to be selective for zinc, not conferring
CC       resistance to cobalt nor cadmium.
CC   -!- SIMILARITY: BELONGS TO THE CATION DIFFUSION FACILITATOR FAMILY OF
CC       TRANSPORTERS (CDF, TC 2.A.4). SLC30A SUBFAMILY.
CC   -----
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   -----
DR   EMBL; AE000177; AAC73839.1; -.
DR   EMBL; D90714; BAA35414.1; -.
DR   EcoGene; EG13662; zitB.
DR   InterPro; IPR002524; Cation_efflux.
DR   Pfam; PF01545; Cation_efflux; 1.
DR   TIGRFAMs; TIGR01297; CDF; 1.
KW   Transport; Zinc transport; Transmembrane; Complete proteome.
FT   TRANSMEM     21     41       POTENTIAL.
FT   TRANSMEM     48     68       POTENTIAL.
FT   TRANSMEM     90    110       POTENTIAL.
FT   TRANSMEM    122    142       POTENTIAL.
FT   TRANSMEM    160    180       POTENTIAL.
FT   TRANSMEM    182    202       POTENTIAL.
FT   DOMAIN      240    313       HIS-RICH.
SQ   SEQUENCE    313 AA;  34678 MW;  36740D67EABAC907 CRC64;

Query Match             22.4%;  Score 431;  DB 1;  Length 313;
Best Local Similarity   32.7%;
Matches  104;  Conservative   69;  Mismatches  121;  Indels   24;  Gaps    9;

Qy      54 HAQKGPDSHCDPKKGKAQORLYVASAICLLFMIGEVVGGYLAHSLAVMIDAAHLLITDFAS 113
           |:   ||||||    ::  : ::::   |:||||||:|||||:||||:::|:::||:|::
Db       3 HSHSHVHSSHL-PEDNNARRLLY-AFGVTAGFMLEVVGGFLSGGSIALLADAGHMLITDTAA 60

Qy     114 MLISLFSLWMSSRPATKTMNFGWQRAEILGALYSVLSISWVVTGVLVYLAVERLISGDYEI 173
           :|  |  || |||||.   | ||.:|   |  ::.: :  :||.||::|.||:|.: :
Db      61 LLFALLAVQFSRRPPIRHTFGWLRLTTLAAFVNATALVVTTILIVWEAIERFRT-PRPV 119

Qy     174 DGGTMLITSGCAVAVNIIMGLTLHQSGHGHSHGTTNQOEENPSVRAAFIHVIGDFMQSMG 233
           :||::.|:: . || ||:  |::|:   |       ||::||||:||:||:|||::.|.|
Db     120 EGGMMMAIAVAGLLANILSFWLLH------HGS---EEKNLNVRAAALHVMGDLLGSVG 169

Qy     234 VLVAAYILYFKPEYKYVDPICTFVFSILVLGTTLFTLRDVILVLMEGTPKGVDFTAV-RD 292
           :|||:::. .   .  .|||:.::|.:||||:|:::|:||.::.:|||:|.:.:|.: .
Db     170 AIIAALIIIW-TGWTPADPILSIVSVLLVLRSAWRLLKDSVNELLEGAPVSLDIAELKRR 228

Qy     293 LLLSVEGVEALHSLHIWALTVAQPVLSVHIAIAQNTDAQAVLKTASSRLQGKFHFHTVTI 352
           :|| ||  |.:||:||:|:  :.|:.: .|.: |::.|   |::.::. : | .|:| :|
Db     229 MCREIPEVRNVHVHVHVW-MVGEKPVMTLHVQVIPPHDHDALLDQIHYLMDHYQIEHATI 287
```

RESULT 10

```
Qy  353  QIEDYSEDMKDCQACQGP  370
             |:|  || :|| ||
Db  288  QME-------YQPCHGP   297
```

```
ID   ZITB_ECO57     STANDARD;      PRT;   311 AA.
AC   Q8X400; Q8X3F7;
DT   15-JUN-2002 (Rel. 41, Created)
DT   15-JUN-2002 (Rel. 41, Last sequence update)
DT   15-JUN-2002 (Rel. 41, Last annotation update)
DE   Zinc transporter zitB.
GN   ZITB OR Z0922 OR ECS0780.
OS   Escherichia coli O157:H7.
OC   Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC   Escherichia.
OX   NCBI_TaxID=83334;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=O157:H7 / EDL933 / ATCC 700927;
RX   MEDLINE=21074935; PubMed=11206551;
RA   Perna N.T., Plunkett G. III, Burland V., Mau B., Glasner J.D.,
RA   Rose D.J., Mayhew G.F., Evans P.S., Gregor J., Kirkpatrick H.A.,
RA   Posfai G., Hackett J., Klink S., Boutin A., Shao Y., Miller L.,
RA   Grotbeck E.J., Davis N.W., Lim A., Dimalanta E.T., Potamousis K.,
RA   Apodaca J., Anantharaman T.S., Lin J., Yen G., Schwartz D.C.,
RA   Welch R.A., Blattner F.R.;
RT   "Genome sequence of enterohaemorrhagic Escherichia coli O157:H7.";
RL   Nature 409:529-533(2001).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=O157:H7 / RIMD 0509952;
RX   MEDLINE=21156231; PubMed=11258796;
RA   Hayashi T., Makino K., Ohnishi M., Kurokawa K., Ishii K., Yokoyama K.,
RA   Han C.-G., Ohtsubo E., Nakayama K., Murata T., Tanaka M., Tobe T.,
RA   Iida T., Takami H., Honda T., Sasakawa C., Ogasawara N., Yasunaga T.,
RA   Kuhara S., Shiba T., Hattori M., Shinagawa H.;
RT   "Complete genome sequence of enterohemorrhagic Escherichia coli
RT   O157:H7 and genomic comparison with a laboratory strain K-12.";
RL   DNA Res. 8:11-22(2001).
CC -!- FUNCTION: Involved in zinc efflux across the cytoplasmic membrane,
CC     thus reducing zinc accumulation in the cytoplasm and rendering
CC     bacteria more resistant to zinc. It may contribute to zinc
CC     homeostasis at low concentrations of zinc (By similarity).
CC -!- SUBCELLULAR LOCATION: Integral membrane protein (Potential).
CC -!- SIMILARITY: BELONGS TO THE CATION DIFFUSION FACILITATOR FAMILY OF
CC     TRANSPORTERS (CDF, TC 2.A.4). SLC30A SUBFAMILY.
CC   --------------------------------------------------------------------------
```

```
DR   EMBL; AE005253; AAG55081.1; -.
DR   EMBL; AP002553; BAB34203.1; -.
KW   Transport; Zinc transport; Transmembrane; Complete proteome.
FT   TRANSMEM     19     39       POTENTIAL.
FT   TRANSMEM     46     66       POTENTIAL.
FT   TRANSMEM     88    108       POTENTIAL.
FT   TRANSMEM    120    140       POTENTIAL.
FT   TRANSMEM    158    178       POTENTIAL.
FT   TRANSMEM    180    200       POTENTIAL.
FT   TRANSMEM    238    311       POTENTIAL.
FT   DOMAIN      238    311       HIS-RICH.
FT   CONFLICT      1      1       MA -> MAHS (IN REF. 2).
SQ   SEQUENCE   311 AA;  34468 MW;  2D0F139E0C537587 CRC64;
```

```
Query Match             22.2%;
Best Local Similarity   32.5%;
                        Score 427; DB 1; Length 311;
                        Pred. No. 9.9e-25;
Matches  104;  Conservative   69;  Mismatches  119;  Indels   28;  Gaps   10;
```

RESULT 11

```
Qy   52  HCHAQKGPDSHCDPRKKGKAQRQLIVASAICLLFMIGEVVGGYLAHSLAVWTDAAHLLTDF  111
             |  ||| | |||   : ||||| ||| :||||||||:|||:|||||||||||||| | |||
Db    3  HSHT----SSHL-PEDNNARRLLY-AFGVTAGFMLVEVIGGFGLSGSLALLADAGHMLTDT   56

Qy  112  ASMLISLFSLWMSSRPATKTMNFGWQRAEILGALVSVLSIMWVTGVLVYLAVERLISGDY  171
             |:| ||| ||  | |||  || ||  :|| ||:| | ||:  ||:|  ||:| || ||||
Db   57  AALLFALLAVQGFSRPPTIRHTFGWLRLTTLAAPVNAIALVVITILIVWEAIERFHT-PR  115

Qy  172  EIDGGTMLITSGCAVAVNIIMGLTLHQSGHGHSHGTTNQQEENPSVRAAFIHVIGDFMQS  231
             :|:||  | || : | ||: ::|| ||:|:|:|  | ||     :|||| ||:|||:  :|
Db  116  PVEGGMMMAIAVAGLLANILSFWLLH------HGS--EEKNLNVRAAALHVLGDLLGS    165

Qy  232  MGVLVAAYILYFKPEYKYDPICTFVFSIIVLLGTTLTIIRDVILVLMEGTPKGVDFTAV-  290
             :| :|||:|| : :| | | :: :|:|:|| | | |:|:| |::||:| ||| ||:|
Db  166  VGAIIAALIIIW-TGWTPADPILSILVLSILVLRSAWRLLKDSVNELLEGAPVSLIDIAELK  224

Qy  291  RDLLLSVEGVEALHSLHIWALTVAQPVLSVHIAIAQNTDAQAVLKTASSRLIQGKFHFHTV  350
             | |  :|||  |||| ||| ||  |||| | |:|  | : | | |:|::  | :| | |||
Db  225  RRMCREIPEYRNVHHVHVW-MVGEKPVMTLHVQVTPPHDHDALLDQIQHYLMDIYQIEHA   283

Qy  351  TIQIEDYSEDMKDCQACQGP  370
             |||   :|  || :|| ||
Db  284  TIQME------YQPCHGP   295
```

```
ID   CZCD_ALCEU     STANDARD;      PRT;   316 AA.
AC   P13512;
DT   01-JAN-1990 (Rel. 13, Created)
DT   01-NOV-1997 (Rel. 35, Last sequence update)
DT   15-JUN-2002 (Rel. 41, Last annotation update)
DE   Cobalt-zinc-cadmium resistance protein czcD (Cation efflux system
DE   protein czcD).
GN   CZCD.
OS   Alcaligenes eutrophus (Ralstonia eutropha).
OG   Plasmid pMOL30.
OC   Bacteria; Proteobacteria; beta subdivision; Ralstonia group;
OC   Ralstonia.
OX   NCBI_TaxID=510;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=CH34;
RX   MEDLINE=90017477; PubMed=2678100;
RA   Nies D.H., Nies A., Chu L., Silver S.;
RT   "Expression and nucleotide sequence of a plasmid-determined divalent
RT   cation efflux system from Alcaligenes eutrophus.";
RL   Proc. Natl. Acad. Sci. U.S.A. 86:7351-7355(1989).
RN   [2]
RP   PRELIMINARY SEQUENCE FROM N.A.
RC   STRAIN=CH34;
RX   MEDLINE=97197194; PubMed=9044283;
RA   van der Lelie D., Schwuchow T., Schwidetzky T., Wuertz S.,
RA   Baeyens W., Mergeay M., Nies D.H.;
RT   "Two-component regulatory system involved in transcriptional control
RT   of heavy-metal homoeostasis in Alcaligenes eutrophus.";
RL   Mol. Microbiol. 23:493-503(1997).
CC -!- FUNCTION: NECESSARY FOR ACTIVATION OF THE CZC DETERMINANT.
CC -!- INDUCTION: BY ZINC.
CC -!- SUBCELLULAR LOCATION: Integral membrane protein (Potential).
CC -!- SIMILARITY: BELONGS TO THE CATION DIFFUSION FACILITATOR FAMILY OF
CC     TRANSPORTERS (CDF, TC 2.A.4). SLC30A SUBFAMILY.
CC   --------------------------------------------------------------------------
```

```
DR   EMBL; X98451; CAA67085.1; -.
```

Left column (RESULT 12)

```
DR   PIR; D33830; D33830.
DR   InterPro; IPR002524; Cation_efflux.
DR   Pfam; PF01545; Cation_efflux; 1.
DR   TIGRFAMs; TIGR01297; CDF; 1.
KW   Plasmid; Transport; Transmembrane; Zinc transport; Cadmium; Cobalt.
FT   TRANSMEM     17     37       POTENTIAL.
FT   TRANSMEM     47     67       POTENTIAL.
FT   TRANSMEM     82    102       POTENTIAL.
FT   TRANSMEM    115    135       POTENTIAL.
FT   TRANSMEM    152    172       POTENTIAL.
FT   TRANSMEM    174    194       POTENTIAL.
SQ   SEQUENCE   316 AA;  33707 MW;  2977C1B4AEE9600F CRC64;

Query Match             20.2%;  Score 388;  DB 1;  Length 316;
Best Local Similarity   34.3%;  Pred. No. 7.8e-22;
Matches  102;  Conservative   57;  Mismatches  120;  Indels  18;  Gaps  7;

Qy   61 SHCDPKKGKAQRQLYVASAICLLFMIGEVVGGYLAHSLAVMTDAAHLLTDFASMLISLFS 120
         |   | :||  | :|  :  ||| :| ||||| |  ||| | |||:||||:||| | :
Db    6 SHDHP--GGNERSLKIALALTGTFLIAEVVGGVWTKSLALISDAAHMLTDTVALAIALAA  63

Qy  121 LWMSSRPATKTMNFGWQRAEILGALVSVLSIWVVTGVLVYLAVERLISGDYEIDGGTMLI 180
         :  |||    |  || :||||||| | ||  :  | :|  | |  |  || :  |||:
Db   64 IAIAKRPADKKRTFGYYRFEILAAAFNALLLFGVAIYILYEAYLRLKSPP-QIESTGMFV 122

Qy  181 TSGCAVAVNIMGLTLHQSGHGHSHGTTNQQEENPSVRAAFIHVIGDFMQSMGVLVAAYI 240
         : |  | | || :|:|:|    |  |  |:|  |   |:|  |  ||: |:|| |  | |
Db  123 VAVLGLIINLI-SMRMLSSG---------OSSSLNVKGAYLEVWSDLLGSVGVIAGAII 171

Qy  241 LYFKPEYKYVDPICTFVFSILVLGTTLTILRDVILVLMEGTPKGVDFTAVRDLLLSVEGV 300
         : | ||  |   | | |: |  |  || :|  ||:||:|| | ||| |  |:||: | |
Db  172 IRF-TGWAWVDSAIAVLIGLWVLPRTWILLKSSLNVLLEGVPDDVDLAEVEKQILATPGV 230

Qy  301 EALHSLHIWALTVAQPVLSVHIA--IAQNTDAQAVLKTASSRLQGKFHFHTVTIQIE 355
         ::| |||||| : ||: : ||:|   ||||  |  | :  :|:|| | |||||| |
Db  231 KSFHDLHIWALTSGKASLTVHVVNDTAVNPEME-VLPELKQMLADKFDITHVTIQFE 286


RESULT 12

ID   C2CD_ALCSP    STANDARD;    PRT;   316 AA.
AC   P94178;
DT   01-NOV-1997 (Rel. 35, Created)
DT   01-NOV-1997 (Rel. 35, Last sequence update)
DT   15-JUN-2002 (Rel. 41, Last annotation update)
DE   Cation efflux system protein czcD.
GN   C2CD.
OS   Alcaligenes sp. (strain CT14).
OC   Bacteria; Proteobacteria; beta subdivision; Alcaligenaceae;
OC   Alcaligenes.
OX   NCBI_TaxID=512;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=96219090; PubMed=8829543;
RA   Kunito T., Kusano T., Oyaizu H., Senoo K., Kanazawa S.,
RA   Matsumoto S.;
RT   "Cloning and sequence analysis of czc genes in Alcaligenes sp. strain
RT   CT14.";
RL   Biosci. Biotechnol. Biochem. 60:699-704(1996).
CC   -!- FUNCTION: NECESSARY FOR ACTIVATION OF THE CZC DETERMINANT.
CC   -!- SUBCELLULAR LOCATION: Integral membrane protein (Potential).
CC   -!- SIMILARITY: BELONGS TO THE CATION DIFFUSION FACILITATOR FAMILY OF
CC       TRANSPORTERS (CDF, TC 2.A.4). SLC30A SUBFAMILY.
```

```
DR   EMBL; D67044; BAA11062.1; -.
```

Right column (RESULT 13)

```
DR   InterPro; IPR002524; Cation_efflux.
DR   Pfam; PF01545; Cation_efflux; 1.
DR   TIGRFAMs; TIGR01297; CDF; 1.
KW   Transport; Transmembrane; Zinc transport; Cadmium; Cobalt.
FT   TRANSMEM     17     37       POTENTIAL.
FT   TRANSMEM     47     67       POTENTIAL.
FT   TRANSMEM     82    102       POTENTIAL.
FT   TRANSMEM    115    135       POTENTIAL.
FT   TRANSMEM    152    172       POTENTIAL.
FT   TRANSMEM    174    194       POTENTIAL.
SQ   SEQUENCE   316 AA;  33741 MW;  D677C1B4A1E964462 CRC64;

Query Match             20.1%;  Score 387;  DB 1;  Length 316;
Best Local Similarity   34.3%;  Pred. No. 9.3e-22;
Matches  102;  Conservative   57;  Mismatches  120;  Indels  18;  Gaps  7;

Qy   61 SHCDPKKGKAQRQLYVASAICLLFMIGEVVGGYLAHSLAVMTDAAHLLTDFASMLISLFS 120
         |   | :||  | :|  :  ||| :| ||||| |  ||| | |||:||||:||| | :
Db    6 SHDHP--GGNERSLKIALALTGTFLIAEVVGGVWTKSLALISDAAHMLTDTVALAIALAA  63

Qy  121 LWMSSRPATKTMNFGWQRAEILGALVSVLSIWVVTGVLVYLAVERLISGDYEIDGGTMLI 180
         :  |||    |  || :||||||| | ||  :  | :|  | |  |  || :  |||:
Db   64 IAIAKRPADKKRTFGYYRFEILAAAFNALLLFGVAIYILYEAYLRLKSPP-QIESTGMFV 122

Qy  181 TSGCAVAVNIMGLTLHQSGHGHSHGTTNQQEENPSVRAAFIHVIGDFMQSMGVLVAAYI 240
         : |  | | || :|:|:|    |  |  |:|  |   |:|  |  ||: |:|| |  | |
Db  123 VAVLGLIINLI-SMRMLSSG---------OSSSLNVKGAYLEVWSDLLGSVGVIAGAII 171

Qy  241 LYFKPEYKYVDPICTFVFSILVLGTTLTILRDVILVLMEGTPKGVDFTAVRDLLLSVEGV 300
         : | ||  |   | | |: |  |  || :|  ||:||:|| | ||| |  |:||: | |
Db  172 IRF-TGWAWVDSAIAVLIGLWVLPRTWFLLKSSLNVLLEGVPDDVDLAEVEKQILATPGV 230

Qy  301 EALHSLHIWALTVAQPVLSVHIA--IAQNTDAQAVLKTASSRLQGKFHFHTVTIQIE 355
         ::| |||||| : ||: : ||:|   ||||  |  | :  :|:|| | |||||| |
Db  231 KSFHDLHIWALTSGKASLTVHVVNDTAVNPEME-VLPELKQMLADKFDITHVTIQFE 286


RESULT 13

ID   ZITB_YERPE    STANDARD;    PRT;   312 AA.
AC   Q8ZGY6;
DT   15-JUN-2002 (Rel. 41, Created)
DT   15-JUN-2002 (Rel. 41, Last sequence update)
DT   15-JUN-2002 (Rel. 41, Last annotation update)
DE   Zinc transporter zitB.
GN   ZITB OR YPO1129.
OS   Yersinia pestis.
OC   Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC   Yersinia.
OX   NCBI_TaxID=632;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=CO-92 / Biovar Orientalis;
RX   MEDLINE=21470413; PubMed=11586360;
RA   Parkhill J., Wren B.W., Thomson N.R., Titball R.W., Holden M.T.G.,
RA   Prentice M.B., Sebaihia M., James K.D., Churcher C., Mungall K.L.,
RA   Baker S., Basham D., Bentley S.D., Brooks K., Cerdeno-Tarraga A.M.,
RA   Chillingworth T., Cronin A., Davies R.M., Davis P., Dougan G.,
RA   Feltwell T., Hamlin N., Holroyd S., Jagels K., Karlyshev A.V.,
RA   Leather S., Moule S., Oyston P.C.F., Quail M., Rutherford K.,
RA   Simmonds M., Skelton J., Stevens K., Whitehead S., Barrell B.G.;
RT   "Genome sequence of Yersinia pestis, the causative agent of plague.";
RL   Nature 413:523-527(2001).
CC   -!- FUNCTION: Involved in zinc efflux across the cytoplasmic membrane,
CC       thus reducing zinc accumulation in the cytoplasm and rendering
CC       bacteria more resistant to zinc. It may contribute to zinc
CC       homeostasis at low concentrations of zinc (By similarity).
CC   -!- SUBCELLULAR LOCATION: Integral membrane protein (Potential).
CC   -!- SIMILARITY: BELONGS TO THE CATION DIFFUSION FACILITATOR FAMILY OF
CC       TRANSPORTERS (CDF, TC 2.A.4). SLC30A SUBFAMILY.
```

```
CC    the European Bioinformatics Institute. There are no restrictions on its
CC    use by non-profit institutions as long as its content is in no way
CC    modified and this statement is not removed. Usage by and for commercial
CC    entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC    or send an email to license@isb-sib.ch).
CC    -------------------------------------------------------------------------
DR    EMBL; AJ141146; CAC89972.1; -.
DR    TIGRFAMs; TIGR01297; CDF; 1.
KW    Transport; Zinc transport; Transmembrane; Complete proteome.
FT    TRANSMEM     16     36       POTENTIAL.
FT    TRANSMEM     40     60       POTENTIAL.
FT    TRANSMEM     81    101       POTENTIAL.
FT    TRANSMEM    117    137       POTENTIAL.
FT    TRANSMEM    153    173       POTENTIAL.
FT    TRANSMEM    177    197       POTENTIAL.
FT    DOMAIN      235    312       HIS-RICH.
SQ    SEQUENCE    312 AA;  34827 MW;  B72E66194A66705 CRC64;

Query Match             19.8%;  Score 380.5;  DB 1;  Length 312;
Best Local Similarity   30.4%;  Pred. No. 2.8e-21;
Matches  92;  Conservative  71;  Mismatches 125;  Indels  15;  Gaps  6;

Qy     72 RQLYVASAICLLFMIGEVVGGYLAHSLAVMTDAAHLLTDFASMLISLFSLWMSSRPATKT 131
          :|| ||   | ||||| |: |  ||:|:|| |||| : | ||| | |||:| :|  |
Db     14 KRLLIAFAITTLFMVTEAIGGWLSGSLALLADAGHMLTDSAALFIALMAVHFSQRKPDPR  73

Qy    132 MNFGWQRAEILGALVSVLSIWVTGVLVYLAVERLISGDYEIDGGTMLITSGCAVAVNII 191
          || |:|||||| ||::|| |   || | |:  || |:| ||:| ||||  |:| :|||
Db     74 HTFGYLRLTLAAFVNAAALLIVILIVWEAVHRFFS-PHEVMGTPMLIIAIAGLLANIF 132

Qy    192 MGLTLHQSGHGHSHGTTNQQEENPSVRAAFIHVIGDFMQSMGVLVAAYILYFKPEYKYVD 251
          ||: |||  || |     |    || ||| | : |  |: || : |:||  : |:
Db    133 CFWILHK------GEEEKNINVRAAALHVLSDLLGSVGAMIAA-IVILTGWTPID 181

Qy    252 PICTEVFSILVLGTTLTILILRDVILVLMEGTPKGVDFTAVR-DLLLSVEGVEALHSLHIWA 310
          ||  |||  |  ||:  || |:|| ||| |: || : | || |||| |||| :| |||
Db    182 PILSVLVSLLIRSAWRLLKESFHELLEGAPQEIDINKLRKDLCTNIYEVRNIHHVRLWQ 241

Qy    311 LTVAQPVLSVHIAIAQNTDAQAVLKTASSRLQGKFHFHTVTIQIEDYSEDMKDCQACQ-G 369
          | ||        |:||||||||||||||||||| ::  |   | :
Db    242 VG-EQRLMTLHAQVIPPLDHDALLQRIQDYLLHHYRISHATVQMEYCQHCGTPDCGINQAA 300

Qy    370 PAD 372
          |:|
Db    301 PAD 303
```

```
RESULT 14
ZRC1_YEAST

ID   ZRC1_YEAST              STANDARD;      PRT;    442 AA.
AC   P20107;
DT   01-FEB-1991 (Rel. 17, Created)
DT   01-FEB-1996 (Rel. 33, Last sequence update)
DT   15-JUN-2002 (Rel. 41, Last annotation update)
DE   Zinc/cadmium resistance protein.
GN   ZRC1 OR YMR243C OR YM9408.05C.
OS   Saccharomyces cerevisiae (Baker's yeast).
OC   Eukaryota; Fungi; Ascomycota; Saccharomyceta; Saccharomycotina;
OC   Saccharomycetes; Saccharomycetales; Saccharomycetaceae; Saccharomyces.
OX   NCBI_TaxID=4932;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=90136503; PubMed=2693940;
RA   Kamizono A., Nishizawa M., Teranishi Y., Murata K., Kimura A.;
RT   "Identification of a gene conferring resistance to zinc and cadmium
RT   ions in the yeast Saccharomyces cerevisiae.";
RL   Mol. Gen. Genet. 219:161-167(1989).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=S288c / AB972;
RA   Gentles S., Barrell B.G., Rajandream M.A., Walsh S.V.;
RL   Submitted (MAR-1995) to the EMBL/GenBank/DDBJ databases.
```

```
-!- FUNCTION: PROBABLY RESPONSIBLE FOR THE UPTAKE OF ZINC AND CADMIUM
    IONS.
-!- SUBCELLULAR LOCATION: Integral membrane protein. Mitochondrial
    (Potential).
-!- SIMILARITY: BELONGS TO THE CATION DIFFUSION FACILITATOR FAMILY OF
    TRANSPORTERS (CDF; TC 2.A.4). SLC30A SUBFAMILY.
--------------------------------------------------------------------------
This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).
--------------------------------------------------------------------------
DR   EMBL; X17537; CAB56542.1; -.
DR   EMBL; Z48756; CAA88653.1; -.
DR   PIR; JQ0349; JQ0349.
SGD; S0004856; ZRC1.
InterPro; IPR002524; Cation_efflux.
Pfam; PF01545; Cation_efflux; 1.
TIGRFAMs; TIGR01297; CDF; 1.
KW   Transport; Transmembrane; Glycoprotein; Zinc transport; Cadmium;
KW   Mitochondrion.
FT   TRANSMEM     10     26       POTENTIAL.
FT   TRANSMEM     42     58       POTENTIAL.
FT   TRANSMEM     80     96       POTENTIAL.
FT   TRANSMEM    113    129       POTENTIAL.
FT   TRANSMEM    240    256       POTENTIAL.
FT   TRANSMEM    270    286       POTENTIAL.
FT   DOMAIN      141    145       HIS-RICH; COULD BE INVOLVED IN
FT                                COORDINATION OF ZINC OR CADMIUM IONS.
FT   DOMAIN      163    167       HIS-RICH; COULD BE INVOLVED IN
FT                                COORDINATION OF ZINC OR CADMIUM IONS.
FT   DOMAIN      216    220       HIS-RICH; COULD BE INVOLVED IN
FT                                COORDINATION OF ZINC OR CADMIUM IONS.
FT   CONFLICT    414    414       R -> S (IN REF. 1).
FT   CONFLICT    417    417       T -> I (IN REF. 1).
SQ   SEQUENCE    442 AA;  48344 MW;  CBE6A9716C0CEBE CRC64;

Query Match             18.6%;  Score 357;  DB 1;  Length 442;
Best Local Similarity   25.2%;  Pred. No. 2.0e-19;
Matches  97;  Conservative  80;  Mismatches 108;  Indels 100;  Gaps  13;

Qy     72 RQLYVASAICL--LFMIGEVVGGYLAHSLAVMTDAAHLLTDFASMLISLFSLWMS-SRPA 128
          :||  || |    | |:|:||  |||   |  | ||:  |||||:| :|||:| :||
Db      5 KELRIISLLTLDIVFFLLEITICGYMSHSLALIADSFHMLNDIISLLVALMAVDVAKNRGP  64

Qy    129 TKTMNFGWQRAEILGALVSVLSIWVTGVLVLYLAVERLISGDYEIDGGTMLITSGCAVAV 188
          :|::|||||:||||||  :  | |   ||:| :   | | : :||:|| ||||  |:|
Db     65 DAKYTYGWKRAEILGALINAVFLIALCFSIMIEALQRLIE-PQEIQNPRLVLYVGVAGLI 123

Qy    189 NIIMGLTL-HQSG---HGHSHGTTN-------------------------- 209
          : ||:| |:||   :|||:||||
Db    124 SNVVGLFLFHDHGSDSLHSHSHGSVESGGNNDLDIESNATHSHSHASLPNDNLAIDEDAIS 183

Qy    210 -------------------QQEENP--------------- 210
          |:||||
Db    184 SPGPSGQIGEVPLPQSVVNRLSNESQPLLNHDDHDHSHESKKPGHRSLNMHGVFLHVLGDA 243

Qy    229 MQSMGVLVAAYILYFKPEYK---YVDPICTEVFSILVLGTTLTILILRDVILVLMEGTPKGV 285
          |:|:|| | |||| |:| |   ||| |: :  :| |||:|| || :||:|| || |: ||
Db    244 LGNIGVIAAALFIW-KTEYSWRYYSDPIVSLIITIIFFSSALPLSRASRILLQATPSTI 302

Qy    286 DFTAVRDLLLSVEGVEALHSLHIWALTVAQPVLSVHIAIAQNTDAQAVLKTASSRLQGK- 344
          |::| | | |:|| |||||:|||||::  ||:| |:|: |:   |: :|||| || |
Db    303 SADQIQREILAVPGVIAVHDFHVWNLTESIYIASIHVQI---DCAPDKFMSSAKLIRKI 358

Qy    345 FH---FHIVTIQIE----DYSEDMK 362
          |:   |   | ||     || :|:
Db    359 FHQHGIHSATVQPEFVSGDVNEDIR 383
```

RESULT 15

ID   COT1_YEAST   STANDARD;   PRT;   439 AA.
AC   P32798;
DT   01-OCT-1993 (Rel. 27, Created)
DT   01-OCT-1996 (Rel. 34, Last sequence update)
DT   30-MAY-2000 (Rel. 39, Last annotation update)
DE   Cobalt uptake protein COT1.
GN   COT1 OR YOR316C OR O6131.
OS   Saccharomyces cerevisiae (Baker's yeast).
OC   Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC   Saccharomycetales; Saccharomycetaceae; Saccharomyces.
OX   NCBI_TaxID=4932;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=92375034; PubMed=1508175;
RA   Conklin D.S., McMaster J.A., Culbertson M.R., Kung C.;
RT   "COT1, a gene involved in cobalt accumulation in Saccharomyces
RT   cerevisiae.";
RL   Mol. Cell. Biol. 12:3678-3688(1992).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=S288c / FY1679;
RX   MEDLINE=97051589; PubMed=8896266;
RA   Pearson B.M., Hernando Y., Payne J., Wolf S.S., Kalogeropoulos A.,
RA   Schweizer M.;
RT   "Sequencing of a 35.71 kb DNA segment on the right arm of yeast
RT   chromosome XV reveals regions of similarity to chromosomes I and
RT   XIII.";
RL   Yeast 12:1021-1031(1996).
CC   -!- FUNCTION: PROBABLY RESPONSIBLE FOR THE UPTAKE OF COBALT IONS. IT
CC   APPEARS TO ACT IN A DOSAGE-DEPENDENT MANNER TO COUNTERACT THE
CC   ADVERSE EFFECTS OF COBALT IONS ON CELLS. IT MAY PARTICIPATE IN
CC   THE REGULATION OF COBALT LEVELS UNDER NORMAL PHYSIOLOGICAL
CC   CONDITIONS AND MAY BE IMPORTANT IN THE SUPPLY OF METAL THAT IS
CC   REQUIRED FOR METALLOENZYME OR COFACTOR SYNTHESIS. IT REDUCES THE
CC   TOXICITY OF COBALT AND RHODIUM IONS. OTHER COMPONENTS RESPONSIBLE
CC   FOR COBALT TRANSPORT EXIST.
CC   -!- SUBCELLULAR LOCATION: MITOCHONDRIAL MEMBRANE. ANOTHER POSSIBILITY
CC   EXISTS THAT IT IS ASSOCIATED WITH ANOTHER UNIDENTIFIED MEMBRANE
CC   THAT HAS BEEN ENRICHED IN THE MITOCHONDRIAL MEMBRANE FRACTIONS.
CC   -!- SIMILARITY: BELONGS TO THE CATION DIFFUSION FACILITATOR FAMILY OF
CC   TRANSPORTERS (CDF, TC 2.A.4). SLC30A SUBFAMILY.
CC   --
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   --
DR   EMBL; M88252; AAA74884.1; -.
DR   EMBL; X90565; CAA62171.1; -.
DR   EMBL; Z75224; CAA99636.1; -.
DR   PIR; S31302; S31302.
SGD; S005843; COT1.
DR   InterPro; IPR002524; Cation_efflux.
Pfam; PF01545; Cation_efflux; 1.
DR   TIGRFAMs; TIGR01297; CDF; 1.
KW   Transport; Cobalt; Mitochondrion; Transmembrane.
FT   TRANSMEM   10   27   POTENTIAL.
FT   TRANSMEM   43   60   POTENTIAL.
FT   TRANSMEM   78   100   POTENTIAL.
FT   TRANSMEM   114   133   POTENTIAL.
FT   TRANSMEM   244   265   POTENTIAL.
FT   TRANSMEM   279   295   POTENTIAL.
FT   DOMAIN   140   148   HIS-RICH; COULD BE INVOLVED IN
FT   COORDINATION OF COBALT IONS.
FT   DOMAIN   163   169   HIS-RICH; COULD BE INVOLVED IN
FT   COORDINATION OF COBALT IONS.
FT   DOMAIN   227   227   HIS-RICH; COULD BE INVOLVED IN
FT   COORDINATION OF COBALT IONS.
FT   CONFLICT   333   334   G -> E (IN REF. 1).
FT   CONFLICT   227   227   HI -> RV (IN REF. 1).

FT   CONFLICT   424   424   I -> V (IN REF. 1).
SQ   SEQUENCE   439 AA;   48154 MW; AC88AAA5F2EE4AED CRC64;

Query Match   17.5%;   Score 337; DB 1; Length 439;
Best Local Similarity   23.8%;   Pred. No. 6.6e-18;
Matches   91; Conservative 83; Mismatches 104; Indels 104; Gaps   11;

Qy   72 RQLYASAICL--LFMIGEVVGGYLAHSLAVMTDAAHLLTDFASMLISLFSLWMS-SRPA 128
         :!:. :! :!   :! ::!!: :! ::!!!!:!!:: ! :!! ! :!!:!!.:: .:!
Db   6 KQVKIISLLLLDTVFFGIEITFGYLSHSLAILIADSFHMLNDIISLVVALWAVNAKRNP 65

Qy 129 TKTMNFGWQRAEILGALVSVLSIWWTVGVLVYLAVERLISGDYEIDGGTMLIITSGCAAV 188
         : :!!:!!!!!:!!!:: ! :! : :!:. :! ::!::: ! :! :!:
Db   66 DSTYTYGWKRAEILGALINAVFLIALCVSILIEALQRIIAPPV-IENPKFVLYGVAGLI 124

Qy 189 NIIMGLTL------HQSGHGHSHG------------------------TT------ 208
         : ! : !!   ! !!!!!!!   !!
Db 125 SNTVGLFLFHDNDQEHGHGHSHGGIFADHEMHMPSSHHTHAHVDGIENTTPMDSTDN 184

Qy 209 -----------------------------NQQEENPSVRAA 220
                                        !::!: .:
Db 185 ISEIMPNAIVDSFMNENTRLLTPENASKTPSYSTSSHTIASGGNYTEHHNKRSLNMHGV 244

Qy 221 FIHVIGDFMQSMGVLVAAYILYFKPEYK---YVDPICTFVFSILVLGTTLLTILRDVILVL 277
         !!!!!:! !! : :! :! :! ::! !   !.! !! !! !:!! !!! :!!: !
Db 245 FLHVLGDALGNIGVMLSAFFIW-KIDYSWKYTDPLVSLLIITGIIFFSSALPLSCKASKIL 303

Qy 278 MEGTPKGVDFTAVRDLLLSVEGVEALHSILHIWALTIVAQPVLSVHIAI----AQNTDAQAV 333
         !:! !! ! !: !!::!!:!: !:!! :!! :!! : !   !! !
Db 304 LQATPSTLSGDQVEGDLLKIPGIIAIHDFHIWNLTESIFIASLHIQLDISPEQFTD---L 360

Qy 334 LKTASSRLQGKFHFHTVTICIE 355
         !!::!: : :!::!!:!:!
Db 361 AKIVRSKLH-RYGIHSATLQPE 381

RESULT 16

ID   ZNT1_RAT   STANDARD;   PRT;   507 AA.
AC   Q62720;
DT   30-MAY-2000 (Rel. 39, Created)
DT   30-MAY-2000 (Rel. 39, Last sequence update)
DT   15-JUN-2002 (Rel. 41, Last annotation update)
DE   Zinc transporter 1 (ZnT-1).
GN   SLC30A1 OR ZNT1.
OS   Rattus norvegicus (Rat).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX   NCBI_TaxID=10116;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Kidney;
RX   MEDLINE=95188868; PubMed=7882967;
RA   Palmiter R.D., Findley S.D.;
RT   "Cloning and functional characterization of a mammalian zinc
RT   transporter that confers resistance to zinc.";
RL   EMBO J. 14:639-649(1995).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Intestine;
RX   MEDLINE=99188868; PubMed=9560190;
RA   McMahon R.J., Cousins R.J.;
RT   "Regulation of the zinc transporter ZnT-1 by dietary zinc.";
RL   Proc. Natl. Acad. Sci. U.S.A. 95:4841-4846(1998).
CC   -!- FUNCTION: MAY BE INVOLVED IN ZINC TRANSPORT OUT OF THE CELL.
CC   -!- SUBUNIT: MULTIMER (PROBABLE).
CC   -!- SUBCELLULAR LOCATION: Integral membrane protein.
CC   LOCALIZED ON THE PLASMA MEMBRANE (PROBABLE). LOCALIZED ON THE
CC   BASOLATERAL SURFACE OF THE ENTEROCYTES.
CC   -!- TISSUE SPECIFICITY: WIDELY EXPRESSED. THE PROTEIN IS DETECTED IN
CC   TISSUE SPECIFICITY: WIDELY EXPRESSED. THE PROTEIN IS DETECTED IN
CC   DUODENUM AND JEJUNUM BUT NOT IN ILEUM AND COLON.
CC   -!- INDUCTION: SLIGHTLY BY ZINC IN THE INTESTINE, BUT NOT THE LIVER.
CC   -!- INDUCTION BY ZINC.
CC   -!- SIMILARITY: BELONGS TO THE CATION DIFFUSION FACILITATOR FAMILY OF

```
RESULT 17
ID   ZNT1_MOUSE     STANDARD;      PRT;   503 AA.
AC   Q60738;
```

LEFT COLUMN:

```
CC   TRANSPORTERS (CDF, TC 2.A.4), SLC30A SUBFAMILY.
CC   -----------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   -----------------------------------------------------------------------
DR   EMBL; U17133; AAA79234.1; -.
DR   InterPro; IPR002524; Cation_efflux.
DR   Pfam; PF01545; Cation_efflux; 1.
DR   TIGRFAMs; TIGR01297; Transport; CDF; 1.
KW   Zinc transport; Transport; Transmembrane; Multigene family; Repeat.
FT   DOMAIN        1     10       CYTOPLASMIC (POTENTIAL).
FT   TRANSMEM     11     31       POTENTIAL.
FT   DOMAIN       32     35       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM     36     56       POTENTIAL.
FT   DOMAIN       57     78       CYTOPLASMIC (POTENTIAL).
FT   TRANSMEM     79     99       POTENTIAL.
FT   DOMAIN      100    113       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM    114    134       POTENTIAL.
FT   DOMAIN      135    247       CYTOPLASMIC (POTENTIAL).
FT   TRANSMEM    248    268       POTENTIAL.
FT   DOMAIN      269    307       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM    308    328       POTENTIAL.
FT   DOMAIN      329    507       CYTOPLASMIC (POTENTIAL).
FT   DOMAIN      145    156       6 X 2 AA APPROXIMATE REPEATS OF H-G.
FT   CARBOHYD    298    298       N-LINKED (GLCNAC...) (POTENTIAL).
SQ   SEQUENCE   507 AA;   55142 MW;   9F977001 7C2455FC CRC64;

Query Match             17.4%;   Score 335;   DB 1;   Length 507;
Best Local Similarity   24.0%;   Pred. No. 1.1e-17;
Matches   105;   Conservative   73;   Mismatches   118;   Indels   142;   Gaps   13;

Qy    68 GKAQRQLYVASAICLLFMIGEVVGGYLAHSLAVMTDAAHLLTDFASMLISLFSLWMSSRP 127
          ||: :|       |||    :||  ||| |  |  |:: | ||| | |:|:|  ||| |
Db     5 GRNRGRLLCMLLLTFMFMVLEVVVSRVTASLAMLSDSFHMLSDVLALVVALVAERFARRT 64

Qy   128 -AIKTMNFGWQQRAEILGALVSVLSIWVVTGVLVLYLAVERLISGDYEIDGGTMLITSGCAV 186
          :||:      |:|||:||  ||||  || | |   :|  |:|||||||||||||   |
Db    65 HATQKNTFGWIRAEVMGALVNAIFLTGICFAILLEAVERFIE-PHEMQQPLVVLSVGVAG 123

Qy   187 AVNIIMGLTL--HQS------GHGHSHG------------------------------ 206
          |||          |||
Db   124 LIVNVLGLCLFHHHSGECQGAGHGHSHGHGHGHLAKGARKAGRAGGVEAGAPPGRAPDQEP 183

Qy   207 ---------------TTN------------QQEENP----S 216
                          :|:
Db   184 DQEETNTLVANTSNSNGLKADQAEPEKLRSDDPVDVQVNGNLIQESDSLESEDNRAQQLN 243

Qy   217 VRAAFIHVIGDFMQSMGVLVAAYLIFK--------------------------------P 245
          :|          |
Db   244 MRGVFLHVLGDALGSVTVVVVNALVFFSWKGCTEDDRCVNPCFPDPCKSSVELMNSTQAP 303

Qy   246 EYK---------YVDPICTFVFSIVLGTTLTILRDVILVLMEGTPKGVDFTAVRDLLLSVE 298
                      :|:
Db   304 MHEAGPCWVLYLDPTLCIIMVCILLYTTYPLIKESALILLQTVPKQIDIKIHLVKELRDVE 363

Qy   299 GVEALHSLHIWALTVAQPVLSVHIAIAQNTDAQAVLKTASSRLQGKFH----FHTVTIQIE 355
          ||  |:   |  : :: | |    |:|           |: |  ||     |  |  |
Db   364 GVEEVIEHLHWQLAGSRRIIATPAHIKCEDPASMQVAKT---IKDVPHNGHGHATTIQPE 419

Qy   356 DYSEDMKD----CQ-ACQ 368
               :|    || ||
Db   420 FASVGSKSSVVPCELACR 437
```

RIGHT COLUMN:

```
ID   ZNT1_MOUSE     STANDARD;      PRT;   503 AA.
AC   Q60738;
DT   30-MAY-2000 (Rel. 39, Created)
DT   30-MAY-2000 (Rel. 39, Last sequence update)
DT   15-JUN-2002 (Rel. 41, Last annotation update)
DE   Zinc transporter 1 (ZnT-1).
GN   SLC30A1 OR ZNT1.
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX   NCBI_TaxID=10090;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Brain;
RX   MEDLINE=95188868; PubMed=7822967;
RA   Palmiter R.D., Findley S.D.;
RT   "Cloning and functional characterization of a mammalian zinc
RT   transporter that confers resistance to zinc.";
RL   EMBO J. 14:639-649(1995).
CC   -!- FUNCTION: MAY BE INVOLVED IN ZINC TRANSPORT OUT OF THE CELL.
CC       LETHALITY OF KNOCKOUT EARLY IN GESTATION SUGGESTS A ROLE OF THE
CC       PROTEIN IN FETAL ZINC ACQUISITION AND RETENTION.
CC   -!- SUBUNIT: MULTIMER (PROBABLE).
CC   -!- SUBCELLULAR LOCATION: INTEGRAL membrane protein.
CC       LOCALIZED ON THE PLASMA MEMBRANE (PROBABLE).
CC   -!- TISSUE SPECIFICITY: WIDELY EXPRESSED.
CC   -!- SIMILARITY: BELONGS TO THE CATION DIFFUSION FACILITATOR FAMILY OF
CC       TRANSPORTERS (CDF, TC 2.A.4), SLC30A SUBFAMILY.
CC   -----------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   -----------------------------------------------------------------------
DR   EMBL; U17132; AAA79233.1; -.
DR   MGD; MGI:1134528; Slc30a1.
DR   InterPro; IPR002524; Cation_efflux.
DR   Pfam; PF01545; Cation_efflux; 1.
DR   TIGRFAMs; TIGR01297; CDF; 1.
KW   Zinc transport; Transport; Transmembrane; Multigene family; Repeat.
FT   DOMAIN        1     10       CYTOPLASMIC (POTENTIAL).
FT   TRANSMEM     11     31       POTENTIAL.
FT   DOMAIN       32     35       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM     36     56       POTENTIAL.
FT   DOMAIN       57     78       CYTOPLASMIC (POTENTIAL).
FT   TRANSMEM     79     99       POTENTIAL.
FT   DOMAIN      100    113       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM    114    134       POTENTIAL.
FT   DOMAIN      135    243       CYTOPLASMIC (POTENTIAL).
FT   TRANSMEM    244    264       POTENTIAL.
FT   DOMAIN      265    303       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM    304    324       POTENTIAL.
FT   DOMAIN      325    503       CYTOPLASMIC (POTENTIAL).
FT   DOMAIN      145    156       6 X 2 AA APPROXIMATE REPEATS OF H-G.
FT   CARBOHYD    294    294       N-LINKED (GLCNAC...) (POTENTIAL).
SQ   SEQUENCE   503 AA;   54716 MW;   7C4FF93F C13CDA22 CRC64;

Query Match             17.3%;   Score 333;   DB 1;   Length 503;
Best Local Similarity   23.7%;   Pred. No. 1.5e-17;
Matches   103;   Conservative   72;   Mismatches   121;   Indels   138;   Gaps   12;

Qy    68 GKAQRQLYVASAICLLFMIGEVVGGYLAHSLAVMTDAAHLLTDFASMLISLFSLWMSSRP 127
          ||: :|       |||    :||  ||| |  |  |:: | ||| | |:|:|  ||| |
Db     5 GRNRGRLLCMLLLTFMFMVLEVVVSRVTASLAMLSDSFHMLSDVLALVVALVAERFARRT 64

Qy   128 -AIKTMNFGWQQRAEILGALVSVLSIWVVTGVLVLYLAVERLISGDYEIDGGTMLITSGCAV 186
          :||:      |:|||:||  ||||  || | |   :|  |:|||||||||||||   |
Db    65 HATQKNTFGWIRAEVMGALVNAIFLTGICFAILLEAVERFIE-PHEMQQPLVVLSVGVAG 123

Qy   187 AVNIIMGLTL--HQS------GHGHSHG------------------------------TTNQQE 212
          |||          |||                                         :|:
Db   124 LIVNVLGLCLFHHHSGECQGAGHGHSHGHGHGHLAKGARKAGRAGGVEAGAPPGRAPDQEP 183
```

```
Qy    213 ENP-----------------------------------------------------SVRAA 220
                                                                  :: |
Db    184 TNTLVANTSNSNGLIKADQAPEPKLRSDDPVDVQVNGNLIQESDNLEAEDNRAGQLNMRGV 243

Qy    221 FIHVIGDFMQSMGVLVAAYILYFK------------------------PE----------- 246
          |:|:||||| |::|||||||||||                        ||
Db    244 FLHVLGDALGSVIVVNALVFYFNWKGCTEDDFCTNPCFPDPCKSSVEIINSTQAPMRDA 303

Qy    247 ---YKYVDPICTFVFSILVLGTTLTILRDVILVLMEGTPKGVDFTAVRDLLLSVEGVEA 302
             | ||   ||||||||| ||| |||:|||||||:| |||||||| :|:||| | |
Db    304 GPCWVLLYLDPTLCIIMVCILLYTYPLLKESALILLQTVPKQIDIKIHLVKELRDVDGVEE 363

Qy    303 LHSLHIWALTVAQPVLSVHIAIAQNTDAQAVLKTASSRLQGKFH---FHIVTIQIEDYSE 359
          |::| |:| ||||||||||||||||||||||||||| : |||     |||:|||||||||
Db    364 VHELHVWQLAGSRIATAHIKCEDPASYMQVAKT----IKDVFHNHGIHATTIQPEPASV 419

Qy    360 DMKD-----CQ-ACQ 368
          |:| |     || :||
Db    420 GSKSSVLPCELACR 433
```

RESULT 18

```
ID    YC63_SYNY3              STANDARD;      PRT;   310 AA.
AC    P74068;
DT    01-NOV-1997 (Rel. 35, Created)
DT    01-NOV-1997 (Rel. 35, Last sequence update)
DT    15-JUN-2002 (Rel. 41, Last annotation update)
DE    Hypothetical protein sll1263.
GN    SLL1263.
OS    Synechocystis sp. (strain PCC 6803).
OC    Bacteria; Cyanobacteria; Chroococcales; Synechocystis.
OX    NCBI_TaxID=1148;
RN    [1]
RP    SEQUENCE FROM N.A.
RX    MEDLINE=97061201; PubMed=8905231;
RA    Kaneko T., Sato S., Kotani H., Tanaka A., Asamizu E., Nakamura Y.,
RA    Miyajima N., Hirosawa M., Sugiura M., Sasamoto S., Kimura T.,
RA    Hosouchi T., Matsuno A., Muraki A., Nakazaki N., Naruo K.,
RA    Okumura S., Shimpo S., Takeuchi C., Wada T., Watanabe A.,
RA    Yamada M., Yasuda M., Tabata S.;
RT    "Sequence analysis of the genome of the unicellular cyanobacterium
RT    Synechocystis sp. strain PCC6803. II. Sequence determination of the
RT    entire genome and assignment of potential protein-coding regions.";
RL    DNA Res. 3:109-136(1996).
CC    -!- SUBCELLULAR LOCATION: Integral membrane protein (Potential).
CC    -!- SIMILARITY: BELONGS TO THE UPF0018 FAMILY.
CC    --------------------------------------------------------------------
CC    This SWISS-PROT entry is copyright. It is produced through a collaboration
CC    between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC    the European Bioinformatics Institute. There are no restrictions on its
CC    use by non-profit institutions as long as its content is in no way
CC    modified and this statement is not removed. Usage by and for commercial
CC    entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC    or send an email to license@isb-sib.ch).
CC    --------------------------------------------------------------------
DR    EMBL; D90912; BAA18144.1; -.
DR    InterPro; IPR002524; Cation_efflux.
DR    Pfam; PF01545; Cation_efflux; 1.
DR    TIGRFAMs; TIGR01297; CDF; 1.
KW    Hypothetical protein; Transmembrane; Complete proteome.
FT    TRANSMEM     10     30       POTENTIAL.
FT    TRANSMEM     44     64       POTENTIAL.
FT    TRANSMEM     78     98       POTENTIAL.
FT    TRANSMEM    113    133       POTENTIAL.
FT    TRANSMEM    161    181       POTENTIAL.
SQ    SEQUENCE   310 AA;  33324 MW;  822C5946CBEC7122 CRC64;
```

```
Qy     72 RQLYVASAICLLFMIGEVVGGY-LAHSLAVMTDAAHLLTDFASMLISLFSLWMSSRPATK 130
          | || |  |||||||||||| || | |||| |||| | |||| ||||:||||||| | |
Db      7 RPYAVLSIGAALATMGLKIGAYAITGSVGLLSDALESTVNLASAIVAFWALSLAATPADS 66

Qy    131 TMNFGWQRAEILGALVSVSLSIWVTGVLVYLAVERLISG---DYEIDGGTMLITSGCAVA 187
          ||| || ||| | | | |: ||  || ||| |||| |||    |:|| |||||:|||
Db     67 EHPPGHSKAEYFSSGLEGAFIFVAALGIGYSAVERLLSPRPLDQNALGIALAI---AATA 123

Qy    188 VNIIMGLTLHQSGHGHSHGGTTNQQEENPSVRAAFIHVIGDFMQSMGVLVAAYILYFKPEY 247
          | :| |||  | ||||: |   | | || | ||  ||||||||||||||||||||||||
Db    124 LNGTVAWILWRAG-----------KRLNSIALRADSQHLMTDVWTSVGVVVAV-ALIFVTGW 173

Qy    248 KYVDPICTFVFSILVLGTTLTILRDVILVLMEGTPKGVDFTAVRDLLLSVE--------- 298
          | ||||||||||||||||||||||||||||||||||||||||||||||||
Db    174 EWLDPLIALGVGENVLWTGTHLRETISSLMDQSLPPAQLQAITSCFLPEYEDQGVRFHLL 233

Qy    299 -----GVEALHSLHIWALTVAQPVLSVHIAIAQNTDAQAVLKTA-SSRLQGKFHFHTVTI 352
               || | ||||| |||||||||||||||||||||||||||| | || ||| ||: |
Db    234 QTRQAGSQSFISFHV--------LVPGHWTVQRGHDLCEAIETAIAERITGS--------- 277

Qy    353 QIEDYSEDMKDCQACQGPSD 372
          |:: |: || :||   :| |
Db    278 RVTTHLEPLEDPKSWQHPDE 297
```

RESULT 19

```
ID    YGLB_BACST              STANDARD;      PRT;   108 AA.
AC    P30540;
DT    01-APR-1993 (Rel. 25, Created)
DT    01-APR-1993 (Rel. 25, Last sequence update)
DT    16-OCT-2001 (Rel. 40, Last annotation update)
DE    Hypothetical protein in GLGB 5'region (Fragment).
OS    Bacillus stearothermophilus.
OC    Bacteria; Firmicutes; Bacillales; Geobacillus.
OX    NCBI_TaxID=1422;
RN    [1]
RP    SEQUENCE FROM N.A.
RX    MEDLINE=92079888; PubMed=1745226;
RA    Kiel J.A.K.W., Boels J.M., Beldman G., Venema G.;
RT    "Molecular cloning and nucleotide sequence of the glycogen branching
RT    enzyme gene (glgB) from Bacillus stearothermophilus and expression in
RT    Escherichia coli and Bacillus subtilis.";
RL    Mol. Gen. Genet. 230:136-144(1991).
CC    -!- SUBCELLULAR LOCATION: Integral membrane protein (Potential).
CC    -!- SIMILARITY: TO CATION ALCALIGENES EUTROPHUS EFFLUX SYSTEM PROTEIN
CC        CZCD.
CC    --------------------------------------------------------------------
CC    This SWISS-PROT entry is copyright. It is produced through a collaboration
CC    between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC    the European Bioinformatics Institute. There are no restrictions on its
CC    use by non-profit institutions as long as its content is in no way
CC    modified and this statement is not removed. Usage by and for commercial
CC    entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC    or send an email to license@isb-sib.ch).
CC    --------------------------------------------------------------------
DR    EMBL; M35089; AAA22481.2; ALT_INIT.
DR    InterPro; IPR002524; Cation_efflux.
DR    Pfam; PF01545; Cation_efflux; 1.
KW    Hypothetical protein; Transmembrane.
FT    TRANSMEM     24     44       POTENTIAL.
FT    TRANSMEM     55     75       POTENTIAL.
FT    NON_TER     108    108       POTENTIAL.
SQ    SEQUENCE   108 AA;  12046 MW;  D10FF162610D90B0 CRC64;
```

```
Qy     67 KGKAQRQLYVASAICLLFMIGEVVGGYLAHSLAVMTDAAHLLTDFASMLISLFSLWMSSR 126
          | :::::|::::: ||| |:::::||:::|:||:||||||||||||||||||||||||| |
Db     17 QNQSKKTLWITLVLTLFFTIVEIVGGLLSNSLALLSDSAHMASDVLALGLSMIALYLAMR 76
```

Qy   127 PATKTMNFGWQRAEILGALVSVLSIWVVT 155
            ||:| ||:: ::||::||::|::|: :::
Db   77 PPNHRFTFGYLRFEIITSFLNGLTLAIIS 105

RESULT 20
ID   YEAB_BACSU   STANDARD;   PRT;   290 AA.
AC   P46348; O05001;
DT   01-NOV-1995 (Rel. 32, Created)
DT   01-NOV-1995 (Rel. 32, Last sequence update)
DT   01-NOV-1997 (Rel. 35, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   Hypothetical protein yeaB.
GN   YEAB.
OS   Bacillus subtilis.
OC   Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.
OX   NCBI_TaxID=1423;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=168;
RX   MEDLINE=97124186; PubMed=8969499;
RA   Borriss R., Porwollik S., Schroeter R.;
RT   "The 52 degrees-55 degrees segment of the Bacillus subtilis
RT   chromosome: a region devoted to purine uptake and metabolism, and
RT   containing the genes cotA, gabP and guaA and the pur gene cluster
RT   within a 34960 bp nucleotide sequence.";
RL   Microbiology 142:3027-3031(1996).
RN   [2]
RP   SEQUENCE OF 1-276 FROM N.A.
RC   STRAIN=168;
RX   MEDLINE=97109536; PubMed=8951816;
RA   Person A.E., Wray L.V., Fisher S.H.;
RT   "Expression of the Bacillus subtilis gabP gene is regulated
RT   independently in response to nitrogen and amino acid availability.";
RL   Mol. Microbiol. 22:693-701(1996).
CC   -!- SUBCELLULAR LOCATION: Integral membrane protein (Potential).
CC   -!- SIMILARITY: BELONGS TO THE UPF0018 FAMILY.
CC   --
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   --
DR   EMBL; U51115; AAB62307.1; -.
DR   EMBL; U31756; AAC44640.1; -.
DR   EMBL; Z99107; CAB12451.1; -.
DR   Subtilist; BG13329; YeaB.
DR   InterPro; IPR002524; Cation_efflux.
DR   Pfam; PF01545; Cation_efflux; 1.
DR   TIGRFAMs; TIGR01297; CDF; 1.
KW   Hypothetical protein; Transmembrane; Complete proteome.
FT   TRANSMEM   15   35   POTENTIAL.
FT   TRANSMEM   82   102   POTENTIAL.
FT   TRANSMEM   113   133   POTENTIAL.
FT   TRANSMEM   159   179   POTENTIAL.
FT   TRANSMEM   181   201   POTENTIAL.
SQ   SEQUENCE   290 AA; 31784 MW; 642E20CA9088B99D CRC64;

Query Match   7.4%; Score 143; DB 1; Length 290;
Best Local Similarity   23.1%; Pred. No. 0.00098;
Matches   70; Conservative   56; Mismatches 137; Indels   40; Gaps   11;

Qy   66 KKGKAQRQLYVASAICLLFMIGEVVGGYLAHSLAVMTDAAHLLTDFASMLISLFSLWMSS 125
            |||:| |||||| ::|:: ::|| ::|:|:|||||::|||||:|||: |: |:|::|:||
Db   8 KKGESGALVSIAAYLVLSAI--KLIIGYLFHSEALFADGLNNTTDIIASVAVLIGLRISQ 65

Qy   126 RPATKTMNFGWQRAEILGALVSVLSIWVVTGVLVLVLAVERLISGDYEI-DGGTMLITSGC 184
            :|: ||| :||| | : :|| ||::|:|| :||| || ::|:| ||: :|::|: :|
Db   66 KPPDEDHPYGHFRAETIASLIASFIMMVVGLQVLFSAGESIFSAKQETPDMIAAWTAAGG 125

Qy   185 AVAVNIIMGLTLHQSGHGHSHSGTTNQQEENPSVRAAFIHVIGDFMQSMGVLVAAYILYFK 244
            : |:   : |||::| || ::| :::| :: | ::||:|||| ||::||::| | | |
Db   126 AVLMLIVYRNKRLAKVKVKSQALLAAADNKS--DAFVS-IGTF---IGIVAAQFHL-- 176

Qy   245 PEYKYVDPICTFVFSILVLGTTLTILRDVILVLMEGTPKGVDFTAVRDLLLSVEGVEALH 304
            |:|:: ||| :|::::| ::| || ::|: :: | || : | ::||| ::|:
Db   177 ---AWIDTVTAFVIGLLLICKTAWDIFKESSHSLTDGFDI-KDISAYKQTIEKISGVSRLK 232

Qy   305 SLHIWALTVAQPVLSVHIAIAQNTDAQAVLKTASSRLQGKFHFHTVVIQIE--------- 355
            : | | :: ||:| :: | |||:| ::::|: ::| :| |:|::
Db   233 DIKARYLG-----STVHVDVV--VEVSADLNITES---------HDIANEIERRMKEEHAI 277

Qy   356 DYS 358
            |||
Db   278 DYS 280

RESULT 21
ID   YM43_YEAST   STANDARD;   PRT;   510 AA.
AC   Q03218;
DT   01-NOV-1997 (Rel. 35, Created)
DT   01-NOV-1997 (Rel. 35, Last sequence update)
DT   01-NOV-1997 (Rel. 35, Last annotation update)
DE   Hypothetical 56.2 kDa protein in SIP18-SPT21 intergenic region.
GN   YMR177W OR YM8010.07.
OS   Saccharomyces cerevisiae (Baker's yeast).
OC   Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC   Saccharomycetales; Saccharomycetaceae; Saccharomyces.
OX   NCBI_TaxID=4932;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=S288c / AB972;
RA   Churcher C.M., Barrell B.G., Rajandream M.A., Walsh S.V.;
RL   Submitted (JUN-1995) to the EMBL/Genbank/DDBJ databases.
CC   -!- SUBCELLULAR LOCATION: Integral membrane protein (Potential).
CC   -!- SIMILARITY: STRONG, TO YEAST YPL224C.
CC   --
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   --
DR   EMBL; Z49808; CAA89910.1; -.
DR   SGD; S0004789; YMR177W.
DR   InterPro; IPR002524; Cation_efflux.
DR   Pfam; PF01545; Cation_efflux; 1.
DR   TIGRFAMs; TIGR01297; CDF; 1.
KW   Hypothetical protein; Transmembrane.
FT   TRANSMEM   165   185   POTENTIAL.
FT   TRANSMEM   194   214   POTENTIAL.
FT   TRANSMEM   241   261   POTENTIAL.
FT   TRANSMEM   286   306   POTENTIAL.
FT   TRANSMEM   333   353   POTENTIAL.
FT   TRANSMEM   356   376   POTENTIAL.
SQ   SEQUENCE   510 AA; 56209 MW; F3CC9A230FB5DB87 CRC64;

Query Match   6.9%; Score 132.5; DB 1; Length 510;
Best Local Similarity   21.9%; Pred. No. 0.011;
Matches   51; Conservative   48; Mismatches   89; Indels   45; Gaps   8;

Qy   2 EAKEKQHLLDARPAIRSYTGSLWQEGAGWIPLPRPGDLDQAIELAAQSNHHCHAQKGPDS 61
            : |:| | : :| | | | |: : | | | || ::|: | :
Db   95 ESETEQNDIISLGTIRDYKSSKCEQA-----DKPS----SLNLHSHTHSHSHA 142

Qy   62 HCDP------------KKGKAQRQLYVASAICLLFMIGEVVGGYLAHSLAVMTDAAHLLTDFA 112
            |: | |||:| |||||| ::|:: ::|| ::|:|:||||:
Db   143 AHNPLLVLSTEQIRKNAGVRITWWGLGVNVGIAIGKFFGGIVFHSQALFADAIHAISDMV 202

```
Qy  113 SMILSLFSLWMSSRPATKTMNFGWQRAEILGALVSVLSIWVVTGVLVLYLAVERLISGDYE 172
        |:||:|| :|| : :|:::  |:||:|| : | |:|:| :||:||: || :|  |:| :
Db  203 SDLILTLLSVGLAANKPIADYPYGYGKIETVGSL-AVSTILAMAGISI------------- 248

Qy  173 IDGGTMLITSGCAVAVNIMGL---TLHQSGHGHSHGTTNQOEENPSVRAAFI 222
         : || |:|:|  |       :| ||:|:| |:  |:  || |: :|  |:|
Db  249 --GWSSL----CALVGPVIPHTIIDTIGNLGHAHTY-SEDIIEDVTDINAAWI 294
```

RESULT 22
YIIP_ECOLI

ID YIIP_ECOLI   STANDARD;   PRT;   300 AA.
AC P32159;
DT 01-OCT-1993 (Rel. 27, Created)
DT 01-OCT-1993 (Rel. 27, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Hypothetical protein yiiP.
GN YIIP OR B3915 OR Z5459 OR ECS4840.
OS Escherichia coli, and
OS Escherichia coli O157:H7.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Escherichia.
OX NCBI_TaxID=562, 83334;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=K12 / MG1655;
RX MEDLINE=93347969; PubMed=8346018;
RA Plunkett G. III, Burland V.D., Daniels D.L., Blattner F.R.;
RT "Analysis of the Escherichia coli genome. III. DNA sequence of the
RT region from 87.2 to 89.2 minutes.";
RL Nucleic Acids Res. 21:3391-3398(1993).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=O157:H7 / EDL933 / ATCC 700927;
RX MEDLINE=21074935; PubMed=11206551;
RA Perna N.T., Plunkett G. III, Burland V., Mau B., Glasner J.D.,
RA Rose D.J., Mayhew G.F., Evans P.S., Gregor J., Kirkpatrick H.A.,
RA Posfai G., Hackett J., Klink S., Boutin A., Shao Y., Miller L.,
RA Grotbeck E.J., Davis N.W., Lim A., Dimalanta E.T., Potamousis K.,
RA Apodaca J., Anantharaman T.S., Lin J., Yen G., Schwartz D.C.,
RA Welch R.A., Blattner F.R.;
RT "Genome sequence of enterohaemorrhagic Escherichia coli O157:H7.";
RL Nature 409:529-533(2001).
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=O157:H7 / RIMD 0509952;
RX MEDLINE=21156231; PubMed=11258796;
RA Hayashi T., Makino K., Ohnishi M., Kurokawa K., Ishii K., Yokoyama K.,
RA Han C.-G., Ohtsubo E., Nakayama K., Murata T., Tanaka M., Tobe T.,
RA Iida T., Takami H., Honda T., Sasakawa C., Ogasawara N., Yasunaga T.,
RA Kuhara S., Shiba T., Hattori M., Shinagawa H.;
RT "Complete genome sequence of enterohemorrhagic Escherichia coli
RT O157:H7 and genomic comparison with a laboratory strain K-12.";
RL DNA Res. 8:11-22(2001).
CC -!- SUBCELLULAR LOCATION: Integral membrane protein (Potential).
CC -!- SIMILARITY: BELONGS TO THE UPF0018 FAMILY.
CC --
CC --
DR EMBL; L19201; AAB03047.1; -.
DR EMBL; AE000466; AAC76897.1; -.
DR EMBL; AE005622; AAG59108.1; -.
DR EMBL; AP002567; BAB38263.1; -.
DR PIR; S40858; S40858.
DR EcoGene; EG11873; yiiP.
DR InterPro; IPR002524; Cation_efflux.
DR Pfam; PF01545; Cation_efflux; 1.

DR TIGRFAMs; TIGR01297; CDF; 1.
KW Hypothetical protein; Transmembrane; Complete proteome.
FT TRANSMEM   12   32   POTENTIAL.
FT TRANSMEM   39   59   POTENTIAL.
FT TRANSMEM   82   102   POTENTIAL.
FT TRANSMEM   114   134   POTENTIAL.
FT TRANSMEM   156   176   POTENTIAL.
FT TRANSMEM   178   198   POTENTIAL.
SQ SEQUENCE   300 AA; 32927 MW; A9ECC4D6ED9A750F CRC64;

Query Match   6.8%; Score 130.5; DB 1; Length 300;
Best Local Similarity 20.7%; Pred. No. 0.0086;
Matches   58; Conservative   62; Mismatches 127; Indels   33; Gaps   9;

```
Qy   72 ROLVASAICLLFMIGEVVGGYLAHSLAVMTDAAHLLTDFASMLISLFSLWMSSRPATKT 131
         |:||:|  |||  | | | |::|::   ::|    || |:|: |:| |: ||  |||
Db   11 RAAIAATAMASLLLLIKIFAWWYTGSSVSIIAALVDSLVDIGASLINLLVVRYSLQPADDN 70

Qy  132 MNFGWQRAEILGALVSVLSIWVVTGVLVLYLV---AVERLISGDYEIDGGTMLITSGCAVAV 188
        |:|  ::|:| |||:| ::  || ::| |:|      |::|| :|    |  |::|| |:
Db   71 HSFGHGKAESLAALAQSM---FISGSALFLFLTGIIQHLISPTPMTDPDGYGVITIVALIC 127

Qy  189 NIMGLTLHQSGHGHSHGTTNQOEENPSVRAAFIHVIGDFMQSMGVLVAAYILYFKPEYK 248
        : | ::|::  ||  |:  :|      ::|:    |||:|  | ||: :| |:
Db  128 TII--LVSFQR------WVVRRTQSOAVRADMLHYQSDVMMNGAILLALGLSWY--GWH 176

Qy  249 YVDPICTFVPSILVLGTTLTILRDVILVLMEGTPKGVDFTAVRDLLLSVEGVEALHSL-- 306
        | | |  | ||:|: |:| |: |: :  |:  |  :|  | ||::|:|:| | ||
Db  177 RADALFALGIGIYILYSALRMGYEAVQSLLDRALPDEEROEIIDIVTSWPGVSGAHDLRT 236

Qy  307 -------------HIWALTVAQPVLSVHIAIAQNTDAQAVLK 335
                      :||: :  |: : |::|  :|| :|:
Db  237 ROSGPPRFIQIHL-EMEDSLPLVQAHM-VADQVE-QAILR 273
```

RESULT 23
Y449_METJA

ID Y449_METJA   STANDARD;   PRT;   283 AA.
AC Q57891;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Hypothetical protein MJ0449.
GN MJ0449.
OS Methanococcus jannaschii.
OC Archaea; Euryarchaeota; Methanococci; Methanococcales;
OC Methanocaldococcaceae; Methanocaldococcus.
OX NCBI_TaxID=2190;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=JAL-1 / DSM 2661 / ATCC 43067;
RX MEDLINE=96337999; PubMed=8688087;
RA Bult C.J., White O., Olsen G.J., Zhou L., Fleischmann R.D.,
RA Sutton G.G., Blake J.A., Fitzgerald L.M., Clayton R.A., Gocayne J.D.,
RA Kerlavage A.R., Dougherty B.A., Tomb J.-F., Adams M.D., Reich C.I.,
RA Overbeek R., Kirkness E.F., Weinstock K.G., Merrick J.M., Glodek A.,
RA Scott J.L., Geoghagen N.S.M., Weidman J.F., Fuhrmann J.L., Nguyen D.,
RA Utterback T.R., Kelley J.M., Peterson J.D., Sadow P.W., Hanna M.C.,
RA Cotton M.D., Roberts K.M., Hurst M.A., Kaine B.P., Borodovsky M.,
RA Klenk H.-P., Fraser C.M., Smith H.O., Woese C.R., Venter J.C.;
RT "Complete genome sequence of the methanogenic archaeon, Methanococcus
RT jannaschii.";
RL Science 273:1058-1073(1996).
CC -!- SUBCELLULAR LOCATION: Integral membrane protein (Potential).
CC -!- SIMILARITY: BELONGS TO THE UPF0018 FAMILY.
CC --

```
CC   EMBL; U67496; AAB98438.1; -.
DR   TIGR; MJ0449;
DR   InterPro; IPR002524; Cation_efflux.
DR   Pfam; PF01545; Cation_efflux; 1.
DR   TIGRFAMs; TIGR01297; CDF; 1.
KW   Hypothetical protein; Transmembrane; Complete proteome.
FT   TRANSMEM      8     28       POTENTIAL.
FT   TRANSMEM     38     58       POTENTIAL.
FT   TRANSMEM     73     93       POTENTIAL.
FT   TRANSMEM    100    120       POTENTIAL.
FT   TRANSMEM    175    195       POTENTIAL.
SQ   SEQUENCE    283 AA;  31286 MW;  4F9128C4222365FE CRC64;

Query Match             6.8%;  Pred. No. 0.0088;  DB 1;  Length 283;
Best Local Similarity  19.9%;  Conservative  65;  Mismatches 134;  Indels  30;  Gaps  7;
Matches  57;

Qy    82 LLFMIGEVVGGYLAHSLAVMTDAAHLLTDFASMLISLFSLWMSSRPATKTMNFGWQRAEI 141
         ::  ||: :| |:|||::: : |:  || || |: :||| | |::| | :  | || |:
Db    16 ILLGLIKIIIGVYVSISLLSDGIHSLSDVITSIGIIGVKIASKPPDESHPYGHSREC 75

Qy   142 LGALVSVLSIWVTGVLVLYLAVERLISGD--YEIDGGTMLITSGCAVANVIIMGLTLHQSG 200
         | :  |:|: | |  |  |:||:|:  |   :| |  ::|:||  | |  : | | || |
Db    76 LFSFFIGLALFFTAYEIGKFAVERIIYGEVIEVNA----IMVGVAILSIVVKELMTRYSL 131

Qy   201 HGHSHGTTNQQEENPSVRAAFIHVIGDFMQSMGVLVAAYILYFKPEYKYVDPICTFVFSI 260
         | |   :  :| |  ||  || |:|||:|| |:|| | || | : |:|  |  |:| :|
Db   132 F------VGKKLNSQVLIADAYHHRSDALSSVVLVG--LLLQKFGIIYGDAIAGIIVAL 183

Qy   261 LVLGTLTILRDVILVLMEGTPKGVDFTAVRDLLSVEGVEALHSLHIWALTVAQPY-LS 319
         :: |: :|   ::  :  | |  ||   :| : : |: : :   : |: : | |  |:
Db   184 MIAKVAFDICLTNIDILTGRAPPKKFFELIEKEALNVDKVIGVHD--IKAHVGPRIHVE 241

Qy   320 VHIAIAQNTDAQAVLKTASSRLQGKFHFHTVTIQIEDYSEDMKDCQ 365
         :|: : :   :  :| :::   |    || :|:| ||  ::|    :
Db   242 LHVEVPSNISAK------------------EMHDIEVAVKNRLESLENVE 273

RESULT 24
ID   YQR5_CAEEL      STANDARD;      PRT;     251 AA.
AC   Q09308;
DT   01-NOV-1997 (Rel. 35, Created)
DT   01-NOV-1997 (Rel. 35, Last sequence update)
DT   01-NOV-1997 (Rel. 35, Last annotation update)
DE   Hypothetical 28.6 kDa protein F19C6.5 in chromosome X.
GN   F19C6.5.
OS   Caenorhabditis elegans.
OC   Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
OC   Rhabditidae; Peloderinae; Caenorhabditis.
OX   NCBI_TaxID=6239;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Harris B.;
RL   Submitted (JAN-1995) to the EMBL/GenBank/DDBJ databases.
CC   -!- SUBCELLULAR LOCATION: Integral membrane protein (Potential).
CC   -----------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   -----------------------------------------------------------------------
DR   EMBL; Z48006; CAA88049.1; -.
DR   WormPep; F19C6.5; CE01558.
KW   Hypothetical protein; Transmembrane.
FT   TRANSMEM     56     76       POTENTIAL.
FT   TRANSMEM    104    124       POTENTIAL.
FT   TRANSMEM    184    204       POTENTIAL.
```

```
FT   TRANSMEM    208    228       POTENTIAL.
SQ   SEQUENCE    251 AA;  28619 MW;  D2B0BFE7AB1C7B30 CRC64;

Query Match             5.6%;  Score 107;  DB 1;  Length 251;
Best Local Similarity  21.4%;  Pred. No. 0.39;  Mismatches 98;  Indels  54;  Gaps 11;
Matches  53;  Conservative  43;

Qy    34 PRPGCLDLQAIELAAQSNHHCHAQKGPDSHCDPKHGKAQRQLYVASAICLLFMIGEVVGGY 93
         | |  |:|: | :  |:|| |  |      | |     | | |::|  |  ||| | :  ||
Db    25 PQP--DIPRIR----NDSTHSIQOPESMLE---KLNTYYTLAV--VIAQLTIGV 67

Qy    94 LAHSLAVMTDAAHLLTDFASMLISLFSLWMSSRPATKTMNFGWQRAEILG-ALVSVLSIW 152
         |  ||:| ||| : : |    :  :|                   |  ::  ::  ||: |
Db    68 LGNSLTVLVADATRIFADHLEL------------------FQYDRAATPGKRLTEIIIVG 108

Qy   153 VVIGVLVILAVERLISGD----YEIDGGTMLITSGCAVAV--NIIMGL---TLHQSGH 201
         :  :| ::| : : | |      :::  ::| ||   | :      |    || | :
Db   109 ITNVILFLLFVAFLLTASGRAATMEFDINRLYLSIGTAMAMTANTLQTLCHFRTWQRERS 168

Qy   202 GHSHGTTNQQEENPSVRAAFIHVIGDFMQSMGVLVAAYILYFKPEYKYVDPICTFVFSIL 261
          |||  |:   : :| |  |  ::  | :: |:::      | :    |:  :   :|:|
Db   169 YHSSLYTG-SKRNQLMHGFVYHFTAYF-----VLVSSLIIVNKDYLIADVITTYATSLL 222

Qy   262 VLGTTLTI 269
         :|  : |:
Db   223 ILANISSI 230

RESULT 25
ID   KE4L_CAEEL      STANDARD;      PRT;     515 AA.
AC   Q9XTQ7;
DT   16-OCT-2001 (Rel. 40, Created)
DT   16-OCT-2001 (Rel. 40, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   Hypothetical Ke4-like protein H13N06.5 in chromosome X.
GN   H13N06.5.
OS   Caenorhabditis elegans.
OC   Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
OC   Rhabditidae; Peloderinae; Caenorhabditis.
OX   NCBI_TaxID=6239;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Lennard N.;
RC   STRAIN=Bristol N2;
RL   Submitted (OCT-1997) to the EMBL/GenBank/DDBJ databases.
CC   -!- SUBCELLULAR LOCATION: Integral membrane protein (Probable).
CC   -!- SIMILARITY: BELONGS TO THE KE4/CATSUP FAMILY.
CC   -----------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   -----------------------------------------------------------------------
DR   EMBL; Z99942; CAB17070.1; -.
DR   WormPep; H13N06.5; CE18815.
DR   InterPro; IPR002395; Kiningen.
DR   InterPro; IPR003689; Zn_trnprt_zip.
DR   Pfam; PF02535; Zip; 1.
DR   PRINTS; PR00334; KININGEN.
KW   Hypothetical protein; Transmembrane; Glycoprotein.
FT   TRANSMEM     27     47       POTENTIAL.
FT   TRANSMEM     49     69       POTENTIAL.
FT   TRANSMEM    214    234       POTENTIAL.
FT   TRANSMEM    247    267       POTENTIAL.
FT   TRANSMEM    297    317       POTENTIAL.
FT   TRANSMEM    386    406       POTENTIAL.
FT   TRANSMEM    429    449       POTENTIAL.
FT   TRANSMEM    463    483       POTENTIAL.
FT   DOMAIN       92    182       HIS-RICH.
```

RESULT 26

```
FT   CARBOHYD     237    237   N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD     379    379   N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD     488    488   N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD       7      7   N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ   SEQUENCE     515 AA;  55500 MW;  17D7E854F4E1DAAF CRC64;

Query Match             5.4%;  Score 104.5;  DB 1;  Length 515;
Best Local Similarity  20.8%;  Pred. No. 1.3;
Matches   43;  Conservative   34;  Mismatches   65;  Indels   65;  Gaps    8;

Qy  152 WVVTGVLLVYLAVERLISGDYEIDGGTIMLITSGCAVAVNIIMGLTLHQSGHGSHGTINQQ 211
            :|:|:| ||:|:|
Db  302 WVLGGIIAFLTVEKL------------------VRILRGEDGHGSHGSHGSHGGEKKE 340

Qy  212 ------------EENP----SVR-AAFIHVIGDFMQSMGVLVAAYILYFKPEYKIYVDP 252
                        |||    ||  ||
Db  341 TRKDSKDKVAKKEEKPEKDEQSIKVTAYLNLAADF------------------THNFTDG 383

Qy  253 ICTFVFSILVLCTTLTLTILRDVILVLMEGTPKGVDFTAVRDLLLSVSEGVEALHSLHIWALT 312
        ||||| :||:||:||:|||
Db  384 LA--IGASFIAGTTVGIV-TMTIVLVHEVPH-----EIGDFALLIQSGYSKKRAMLIQLV 435

Qy  313 VAQPVLSVHIAIAQNTDAQAVLKTASS 339
        |:||| :||:|:|
Db  436 TALGALSGCVISLFSADADALADAAAS 462
```

```
DR   EMBL; X82304; CAA57747.1; -.
DR   EMBL; L39205; AAC28748.1; -.
DR   HSSP; P00157; 1BE3.
DR   InterPro; IPR000179; Cyt_b_b6.
DR   Pfam; PF00032; cytochrome_b_C; 1.
DR   Pfam; PF00033; cytochrome_b_N; 1.
DR   PROSITE; PS00190; CYTOCHROME_B_HEME; 1.
DR   PROSITE; PS00192; CYTOCHROME_B_QO; 1.
DR   PROSITE; PS00193; CYTOCHROME_B_QI; 1.
KW   Electron transport; Mitochondrion; Respiratory chain; Transmembrane;
KW   Heme.
FT   METAL         83     83   IRON 1 (HEME B562 AXIAL LIGAND).
FT   METAL         97     97   IRON 2 (HEME B566 AXIAL LIGAND).
FT   METAL        182    182   IRON 1 (HEME B562 AXIAL LIGAND).
FT   METAL        196    196   IRON 2 (HEME B566 AXIAL LIGAND).
SQ   SEQUENCE     379 AA;  42657 MW;  905FF04B79EE7E12 CRC64;
```

```
CYB_PHOHI
ID   CYB_PHOHI     STANDARD;      PRT;   379 AA.
AC   Q35468; Q35466;
DT   01-NOV-1997 (Rel. 35, Created)
DT   01-NOV-1997 (Rel. 35, Last sequence update)
DT   30-MAY-2000 (Rel. 39, Last annotation update)
DE   Cytochrome b.
GN   MTCYB OR COB OR CYTB.
OS   Phoca hispida (Ringed seal) (Pusa hispida).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Carnivora; Pinnipedia; Phocidae; Phoca.
OX   NCBI_TaxID=9718;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=95230701; PubMed=7714914;
RA   Arnason U., Bodin K., Gullberg A., Ledje C., Mouchaty S.;
RT   "A molecular view of pinniped relationships with particular emphasis
RT   on the true seals.";
RL   J. Mol. Evol. 40:78-85(1995).
RN   [2]
RP   SEQUENCE OF 1-134 FROM N.A.
RC   TISSUE=Muscle;
RA   Perry E.A., Carr S.M., Bartlett S.E., Davidson W.S.;
RT   "A phylogenetic perspective on the evolution of reproductive behavior
RT   in pagophilic seals of the Northwest Atlantic as indicated by
RT   mitochondrial DNA sequences.";
RL   J. Mammal. 76:22-31(1995).
CC   -!- FUNCTION: COMPONENT OF THE UBIQUINOL-CYTOCHROME C REDUCTASE
CC       COMPLEX (COMPLEX III OR CYTOCHROME B-C1 COMPLEX), WHICH IS A
CC       RESPIRATORY CHAIN THAT GENERATES AN ELECTROCHEMICAL POTENTIAL
CC       COUPLED TO ATP SYNTHESIS.
CC   -!- COFACTOR: TWO HEME GROUPS (B562 AND B566) WHICH ARE NOT COVALENTLY
CC       BOUND TO THE PROTEIN.
CC   -!- SUBUNIT: THE MAIN SUBUNITS OF COMPLEX B-C1 ARE: CYTOCHROME B,
CC       CYTOCHROME C1 AND THE RIESKE PROTEIN.
CC   -!- SIMILARITY: BELONGS TO THE CYTOCHROME B FAMILY.
```

RESULT 27

```
Query Match             5.3%;  Score 102.5;  DB 1;  Length 379;
Best Local Similarity  20.4%;  Pred. No. 1.3;
Matches   69;  Conservative   44;  Mismatches  109;  Indels  117;  Gaps   17;

Qy   80 ICILFMIGEVVGGVLA-----HSLAVMTDAAHLTDF-----------------ASM-LISLF- 119
Db   97 HVGRGLIYGSYTFTETWNIGIILLFTVMATAFMGYVLPWGGQMSFWGWGATV----ITNLLSA 152

Qy  120 ---SLWMSSRPAIKTMNFG---------------WQRAEILGALVSLSIWWVTGV 157
Db  153 IPYVGTDLVQWIWGGFSVDKATLTRFFAFHFILPFVVLALAVVHLLFLHETGSNNPSGIT 212

Qy  158 LVYLAVE--RLISGDYEIDGGTM------LITSGCAVAVNIIMGLTLHQSGHGSHGTT 208
Db  213 SDSDKIP-----FHPYYTIKDILGALLLILVLTLLVLFSPDLLGDPDNYIPANPLSTPP 266

Qy  209 NQQEENPSVRAAFIHVIGDFMQSMG------VLVAAYILYFKPE-----YKYVDPICT- 255
Db  267 HIKPEWYFLFAYAILRSIPNKLGGVIALVLSILLAIMPLLHTSKQRGMMFRPISQCL-- 324

Qy  256 ----------FVFSIL-----VLGTLTLTILRDVILVLME-GTPKGVDFTAVRDLLLS 296
Db  325 -----------FWLLVADLLTLTWIGGQPVEHPYITIGQ 352

Qy  297 VEGVEALHSLHIWALTV----------AQPVLSVHIAIAQ 326
```

```
NU5C_SYNY3
ID   NU5C_SYNY3    STANDARD;      PRT;   681 AA.
AC   Q55429;
DT   01-NOV-1997 (Rel. 35, Created)
DT   01-NOV-1997 (Rel. 35, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   NADH-plastoquinone oxidoreductase chain 5 (EC 1.6.5.3).
GN   NDHF OR SLR0844.
OS   Synechocystis sp. (strain PCC 6803).
OC   Bacteria; Cyanobacteria; Chroococcales; Synechocystis.
OX   NCBI_TaxID=1148;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=96127529; PubMed=8590279;
RA   Kaneko T., Tanaka A., Sato S., Kotani H., Sazuka T., Miyajima N.,
RA   Sugiura M., Tabata S.;
RT   "Sequence analysis of the genome of the unicellular cyanobacterium
RT   Synechocystis sp. strain PCC6803. I. Sequence features in the 1 Mb
RT   region from map positions 64% to 92% of the genome.";
RL   DNA Res. 2:153-166(1995).
CC   -!- CATALYTIC ACTIVITY: NADH + plastoquinone = NAD(+) + plastoquinol.
CC   -!- SUBCELLULAR LOCATION: Integral membrane protein.
CC   -!- SIMILARITY: TO POLYPEPTIDE 5 OF THE NADH-UBIQUINOL OXIDOREDUCTASE
CC       OF CHLOROPLASTS OR MITOCHONDRIA.
```

DR   EMBL; D64003; BAA10530.1; -.
DR   InterPro; IPR003916; NADHub_oxred5.
DR   InterPro; IPR001750; Oxidored_q1.
DR   InterPro; IPR001516; Oxidored_q1_N.
DR   Pfam; PF00361; oxidored_q1; 1.
DR   Pfam; PF00662; oxidored_q1_N; 1.
DR   PRINTS; PR01434; NADHDHGNASE5.
KW   Oxidoreductase; NAD; Plastoquinone; Transmembrane; Complete proteome.
FT   TRANSMEM   7   27   POTENTIAL.
FT   TRANSMEM   39   59   POTENTIAL.
FT   TRANSMEM   89   109   POTENTIAL.
FT   TRANSMEM   120   140   POTENTIAL.
FT   TRANSMEM   144   164   POTENTIAL.
FT   TRANSMEM   188   208   POTENTIAL.
FT   TRANSMEM   219   239   POTENTIAL.
FT   TRANSMEM   258   278   POTENTIAL.
FT   TRANSMEM   289   309   POTENTIAL.
FT   TRANSMEM   327   347   POTENTIAL.
FT   TRANSMEM   352   372   POTENTIAL.
FT   TRANSMEM   395   415   POTENTIAL.
FT   TRANSMEM   420   440   POTENTIAL.
FT   TRANSMEM   509   529   POTENTIAL.
FT   TRANSMEM   558   578   POTENTIAL.
FT   TRANSMEM   660   680   POTENTIAL.
SQ   SEQUENCE   681 AA; 74400 MW; D11C883B722FDAE CRC64;

Query Match   5.3%;   Score 101.5; DB 1; Length 681;
Best Local Similarity 20.7%;   Pred. No. 2.9;
Matches   52; Conservative   36; Mismatches 112; Indels 51; Gaps   8;

Qy   74 LYVASAICLLFMIGEVVGGYLAHSLAVMTDAAHLLTDFASML---IS----------   117
                :|| || ||:|| | ||: | |: | ||| :|| |:
Db   97 LVIVISVALLVMI--YTDGYMAHDPGYVRFYAYLSLFASSMLGLVISPNLVQVYIFWELV 154

Qy   118 -----LFSLWMSSRPATKIMN--IGWQRAEILGALVSVLSIWVTGVLVLVLAVERLISG 169
                 |:|| ||||||||||:|||||||
Db   155 GMCSYLLIGEWYDRKAAADACQKAEVTNRVGDFGLLLGLILGCLYIWAFGSFDGTIGERLEG 214

Qy   170 DYEIDGGTMLITSG------CAVAVNIMGLTLHQSGHGHSHGTTNQQEENPSVRAAFI 222
                :||| |:||
Db   215 -----------LVSSGVLSGAIAAILAILVFLG-PVAKSAQFPLHVWLPDAMEGPTPISALI 264

Qy   223 HVIGDEMQSMGVLVAAYILYFKPEYKYVDPICTFVRSILVLGTTLTLITLRDVLVLMEGTP 282
                :||| ||||
Db   265 HA--ATMVAAGVFLVARMYPVFEPIPVVMNTIAFTGCFTAFLGATIALTQNDIK 316

Qy   283 KGVDFTAVRDL 293
                :||: :||
Db   317 KGLAYSTISOL 327

RESULT 28
CYB_ORNAN
ID   CYB_ORNAN   STANDARD;   PRT;   379 AA.
AC   Q36461;
DT   01-NOV-1997 (Rel. 35, Created)
DT   01-NOV-1997 (Rel. 35, Last sequence update)
DT   30-MAY-2000 (Rel. 39, Last annotation update)
DE   Cytochrome b.
GN   MTCYB OR COB OR CYTB.
OS   Ornithorhynchus anatinus (Duckbill platypus).
OG   Mitochondrion.
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Monotremata; Ornithorhynchidae; Ornithorhynchus.
OX   NCBI_TaxID=9258;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Heart, and Liver;
RX   MEDLINE=97077300; PubMed=8819867;
RA   Janke A., Gemmell N., Feldmaier-Fuchs G., von Haeseler A.,
RA   Paabo S.;
RT   "The mitochondrial genome of a monotreme -- the platypus
RT   (Ornithorhynchus anatinus)";
RL   J. Mol. Evol. 42:153-159(1996).
CC   -!- FUNCTION: COMPONENT OF THE UBIQUINOL-CYTOCHROME C REDUCTASE
CC   COMPLEX (COMPLEX III OR CYTOCHROME B-C1 COMPLEX), WHICH IS A
CC   RESPIRATORY CHAIN THAT GENERATES AN ELECTROCHEMICAL POTENTIAL
CC   COUPLED TO ATP SYNTHESIS.
CC   -!- COFACTOR: TWO HEME GROUPS (B562 AND B566) WHICH ARE NOT COVALENTLY
CC   BOUND TO THE PROTEIN.
CC   -!- SUBUNIT: THE MAIN SUBUNITS OF COMPLEX B-C1 ARE: CYTOCHROME B,
CC   CYTOCHROME C1 AND THE RIESKE PROTEIN.
CC   -!- SIMILARITY: BELONGS TO THE CYTOCHROME B FAMILY.
DR   EMBL; X83427; CAA58454.1; -.
DR   InterPro; IPR000179; Cyt_b6.
DR   Pfam; PF00032; cytochrome_b_C; 1.
DR   Pfam; PF00033; cytochrome_b_N; 1.
DR   PROSITE; PS00192; CYTOCHROME_B_HEME; 1.
DR   PROSITE; PS00193; CYTOCHROME_B_QO; 1.
KW   Electron transport; Mitochondrion; Respiratory chain; Transmembrane;
KW   Heme.
FT   METAL   83   83   IRON 1 (HEME B562 AXIAL LIGAND).
FT   METAL   97   97   IRON 1 (HEME B566 AXIAL LIGAND).
FT   METAL   182   182   IRON 2 (HEME B562 AXIAL LIGAND).
FT   METAL   196   196   IRON 2 (HEME B566 AXIAL LIGAND).
SQ   SEQUENCE   379 AA; 42577 MW; B753F94E31334A7C CRC64;

Query Match   5.2%;   Score 100.5; DB 1; Length 379;
Best Local Similarity 19.8%;   Pred. No. 1.8;
Matches   65; Conservative   52; Mismatches 108; Indels 103; Gaps 16;

Qy   80 ICLLFMIGEVVGGYLA-----HSLAVMTDAAHLLTDF-ASMLI---------- 119
            :||| |: ||:|| |: |
Db   39 LCLIIQI--LTGLFLAMHYTSDTSTAFSSVAHICRDVNYGWLIRYMHANGASLFFMCIFL 96

Qy   120 ----SLWMSSRPATKTMNFG------------------WQRAEILGALVSVLSIWVTGV 157
                |:|| |:| ||||||||||:|||
Db   97 HIGRGLYYGSYQTETWNIGVLLFTVMATAFVGYVLPWGQMSFWGAT----VITNL 149

Qy   158 L-----VYLAVERLISGDYEIDGGTM------LITSGCAVAVNIMGLTLHQSGHGHSH 205
            | |||:||| |:||
Db   150 LSAIPYIGTTLVEWIWGGFSVDKATLIRFFAFHFILCPVIAALAVVIHLLFLHETGSNNPS 209

Qy   206 GTTNQQEENP-----SVRAAFIHVIGDEMQSMGVLVAAYILYFKPE-----YKYDPI 253
                 | :|| :
Db   210 GLNSDPDKIFHYYSVK----DLVGFFMT---ILVLLTLVLFTPDLLGDPDNYTPANPL 262

Qy   254 CT--------FVFSIL-----VLGTTLTLIRDVILVLM----EGTPKGVDFTAVRD 292
            |:   : |||:| |:   :
Db   263 STPPHIKPEWYFLFAYAILRSIPNKLGGVIALVASILILILVPLLHTSYQRGLAFRPLTQ 322

Qy   293 LLLSVEGVEALHSLHIWALTVAQPVLSV 320
            :| | ||: :|
Db   323 MLFWILVTDLLLTLTWIGGQPVEQPFIII 350

RESULT 29
YGJK_CAEEL
ID   YGJK_CAEEL   STANDARD;   PRT;   404 AA.
AC   Q9XUC4;

Record 1 (left column)

```
DT   16-OCT-2001 (Rel. 40, Created)
DT   16-OCT-2001 (Rel. 40, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   Hypothetical 44.9 kDa protein T28F3.3 in chromosome IV.
GN   T28F3.3
OS   Caenorhabditis elegans.
OC   Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
OC   Rhabditidae; Peloderinae; Caenorhabditis.
OX   NCBI_TaxID=6239;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=Bristol N2;
RA   Mortimore B.;
RL   Submitted (NOV-1996) to the EMBL/GenBank/DDBJ databases.
-!- SUBCELLULAR LOCATION: Integral membrane protein (Probable).
-!- SIMILARITY: BELONGS TO THE KE4/CATSUP FAMILY.
CC   -----------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   -----------------------------------------------------------------------
     EMBL; Z82285; CAB05297.1; -.
DR   WormPep; T28F3.3; CE19807.
DR   InterPro; IPR003689; Zn_trnprt_zip.
     Pfam; PF02535; Zip; 1.
KW   Hypothetical protein; Transmembrane; Glycoprotein.
FT   TRANSMEM     85    105       POTENTIAL.
FT   TRANSMEM    117    137       POTENTIAL.
FT   TRANSMEM    171    191       POTENTIAL.
FT   TRANSMEM    255    275       POTENTIAL.
FT   TRANSMEM    303    323       POTENTIAL.
FT   TRANSMEM    332    352       POTENTIAL.
FT   TRANSMEM    369    389       POTENTIAL.
FT   DOMAIN       19     60       HIS-RICH.
FT   DOMAIN      133    164       HIS-RICH.
FT   DOMAIN      195    219       HIS-RICH.
FT   CARBOHYD     66     66       N-LINKED (GLCNAC...) (POTENTIAL).
FT   CARBOHYD    152    152       N-LINKED (GLCNAC...) (POTENTIAL).
FT   CARBOHYD    248    248       N-LINKED (GLCNAC...) (POTENTIAL).
FT   CARBOHYD    362    362       N-LINKED (GLCNAC...) (POTENTIAL).
SQ   SEQUENCE   404 AA;  44884 MW;  77C40C592B3E709 CRC64;

Query Match             5.2%;
Best Local Similarity  21.2%;
Matches   82;  Conservative   62;  Mismatches  119;  Indels  124;  Gaps  22;

Score 100;  DB 1;  Length 404;

Qy     26 EGAGWIPLPRPGLDLQAIELAAQSNHHCHAQKGPDSHCDPKKGKAQROLYVASAICL---  82
Db     28 ----------HDHHEHDHDHDEQLIRKNHTSHREIQHSRLSTLKVW  84

Qy     83 LFMIGEVVGGYLAH-SLAVMTDAAHLILDFASMLISLFSLWMSSRPAIKTMNFGWQRAEI 141
Db     85 VFSLSAVVGISLAPCTLLFFIPAQHANGPFLKILLA------        FG--AGGL 126

Qy    142 LG-ALVSVL-----                       SIWVTGVLVYLAVER 165
Db    127 LGDALLHIIPHSLSPHDHSHDHDHNHSHKEHDHSHDHSNQLRVGTFVIAGILVFMMVEQ 186

Qy    166 LISGDYEIDGGTMLITSGCCAVAVN--IIMGLTLHQSGHGHSHGTTNQQ--EENPSVRA-A 220
Db    187 LVR----IIKGG------HCHSHENGHIVADEHRHLNEHDHEHSEEKKQQVEGLKDVKASA 237

Qy    221 FIHVIGDPMQSM--GVLVAAYILYFKPEYKYVDPICTFVFSI-LVLG--TTLTIL----- 270
Db    238 YLNLVADFVHNVTDGLAIGA------------SFSAGNTLGWITTLTVLLHELP 279

Qy    271 ---RDVILVLMEGTPKGVDFTAVRDLLLSVEGVEALHSLH--IWALIVAQPVLSVHIAIA 325
Db    280 HEVGDPAILVQSGFSK---YQAIR----LQAVTALGAITGCVFSLLVSNPG----SLN 326
```

Record 2 (right column)

```
Qy    326 QNTDAQAVLKTASSRLQGKFHFHTVYI 352
Db    327 NDADISAIMPFTAG---GFIYIAIVSV 350

RESULT 30
CYB_DROYA
Qy    326 QNTDAQAVLKTASSRLQGKFHFHTVYI 352
Db    327 NDADISAIMPFTAG---GFIYIAIVSV 350

ID   CYB_DROYA      STANDARD;      PRT;   378 AA.
AC   P07704;
DT   01-APR-1988 (Rel. 07, Created)
DT   01-APR-1988 (Rel. 07, Last sequence update)
DT   30-MAY-2000 (Rel. 39, Last annotation update)
DE   Cytochrome b.
GN   COB OR CYTB.
OS   Drosophila yakuba (Fruit fly).
OC   Eukaryota; Metazoa; Arthropoda; Mandibulata; Pancrustacea; Hexapoda;
OC   Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
OC   Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
OG   Mitochondrion.
OX   NCBI_TaxID=7245;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=84221393; PubMed=6328435;
RA   Clary D.O., Wahleithner J.A., Wolstenholme D.R.;
RT   "Sequence and arrangement of the genes for cytochrome b, URF1, URF4L,
RT   URF4, URF5, URF6 and five tRNAs in Drosophila mitochondrial DNA.";
RL   Nucleic Acids Res. 12:3747-3762(1984).
-!- FUNCTION: COMPONENT OF THE UBIQUINOL-CYTOCHROME C REDUCTASE
CC   COMPLEX (COMPLEX III OR CYTOCHROME B-C1 COMPLEX), WHICH IS A
CC   RESPIRATORY CHAIN THAT GENERATES AN ELECTROCHEMICAL POTENTIAL
CC   COUPLED TO ATP SYNTHESIS.
-!- COFACTOR: TWO HEME GROUPS (B562 AND B566) WHICH ARE NOT COVALENTLY
CC   BOUND TO THE PROTEIN.
-!- SUBUNIT: THE MAIN SUBUNITS OF COMPLEX B-C1 ARE: CYTOCHROME B,
CC   CYTOCHROME C1 AND THE RIESKE PROTEIN.
-!- SIMILARITY: BELONGS TO THE CYTOCHROME B FAMILY.
CC   -----------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   -----------------------------------------------------------------------
     EMBL; X03240; CAA26996.1; -.
DR   PIR; C30020; C30020.
DR   FlyBase; FBgn0013182; Dyak\mt:Cyt-b.
DR   InterPro; IPR000179; Cyt_b_b6.
DR   Pfam; PF00032; cytochrom_B_C; 1.
DR   Pfam; PF00033; cytochrome_B_N; 1.
DR   PROSITE; PS00192; CYTOCHROME_B_HEME; 1.
DR   PROSITE; PS00193; CYTOCHROME_B_QO; 1.
KW   Electron transport; Mitochondrion; Respiratory chain; Transmembrane;
KW   Heme.
FT   METAL        84     84       IRON 1 (HEME B562 AXIAL LIGAND).
FT   METAL        98     98       IRON 2 (HEME B566 AXIAL LIGAND).
FT   METAL       183    183       IRON 1 (HEME B562 AXIAL LIGAND).
FT   METAL       197    197       IRON 2 (HEME B566 AXIAL LIGAND).
SQ   SEQUENCE   378 AA;  43137 MW;  86849IDE01B03IAF CRC64;

Query Match             5.1%;
Best Local Similarity  18.7%;
Matches   65;  Conservative   54;  Mismatches  111;  Indels  118;  Gaps  16;

Score 99;  DB 1;  Length 378;
```

```
Qy   43  IELAAQS-NHHCHAQKGPDSHCDPKRGKAQRQLYVASA-----ICLLFMIGEVV--GGYLA   95
             :|||| : ||   |:                           ||||||||| |   |||||
Db   60  VNLAFYSVNHICR--------DVNYGWLRTLHANGASFFFICIYLHIGRGIYYGSYLF      110

Qy   96  HSLAVWTDAAHLLTDFASMLISLFSLHMSSRPATKTMNFGWQRAEILGALVSLSIWVVT    155
             ||:   |: |                                  ||
Db  111  ---------TPTWLVGVIILFLVMGTAFMGYVLPWGQMSFWGATV---ITNLL          151

Qy  156  GVLVYLAVE--RLISGDYEIDGGTM------LITSGCAVAVNIIMGLTLHQSGHGHSHG    206
             |:|:   :||: ||||:|:|||||      |||||||||||||||||||||||:|:|:|
Db  152  SAIPYLGMDLVQWLWGGFAVDNATLTRFFTFHFILPFIVLAMTIHLLFLHQTGSNNPIG    211

Qy  207  TTNQEENPSVRAAFIHVIGDDMQSMGVLVAAYIL-----------                  241
             :::   |   |:|                ||
Db  212  LNSNIDKIP-----FHPYTFKDIVGFIVMFILISLVLISPNLLGDPDNFIPANDLVT      265

Qy  242  --YFKPEYKYVDPICTFVFSIL------VLGTTLTILRDVILVLMEGTP------KGVDF   287
             | :|:  : | |:|          |:||||||| :| :| :|||||:||       :|||:
Db  266  PAHIQEWYFL-----FAYAILRSIPNKLGGVIALVLSIAILMI--LPFYNLSKFRGIQF   318

Qy  288  TAVRDLLLSVEGVEALHSLHIWALTV----AQPVLSVHIAIAQ                    326
             |:|   | : : ||:|:| ||:||:      :||: ||:|: |
Db  319  YPINQIL-----------FWSMLVTVIILLTWIGARVEEPVVLIGQ                 353
```

RESULT 31

```
ID   CYB_CYSCR    STANDARD;    PRT;    379 AA.
AC   Q34070; Q34069; Q34065;
DT   15-JUL-1999 (Rel. 38, Created)
DT   15-JUL-1999 (Rel. 38, Last sequence update)
DT   30-MAY-2000 (Rel. 39, Last annotation update)
DE   Cytochrome B.
GN   MTCYB OR COB OR CYTB.
OS   Cystophora cristata.
OG   Mitochondrion.
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Carnivora; Pinnipedia; Phocidae; Cystophora.
OX   NCBI_TaxID=39293;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=95230701; PubMed=7714914;
RA   Arnason U., Bodin K., Gullberg A., Ledje C., Mouchaty S.;
RT   "A molecular view of pinniped relationships with particular emphasis
RT   on the true seals.";
RL   J. Mol. Evol. 40:78-85(1995).
RN   [2]
RP   SEQUENCE OF 1-134 FROM N.A.
RA   Carr S.M., Perry E.A.;
RT   "Intra- and interfamilial systematic relationships of phocid seals as
RT   indicated by mitochondrial DNA sequences.";
RT   (In) Dizon A.E., Chivers S.J., Perrin W.F. (eds);
RL   Molecular genetics of marine mammals, pp.277-290, Society for marine
RL   mammalogy, Kansas (1997).
RN   [3]
RP   SEQUENCE OF 51-130 FROM N.A.
RC   TISSUE=Muscle;
RA   Perry E.A., Carr S.M., Bartlett S.E., Davidson W.S.;
RT   "A phylogenetic perspective on the evolution of reproductive behavior
RT   in pagophilic seals of the Northwest Atlantic as indicated by
RT   mitochondrial DNA sequences.";
RL   J. Mammal. 76:22-31(1995).
```

```
DR   EMBL; X82294; CAA57737.1; -.
DR   EMBL; L19124; AAA74098.1; -.
DR   EMBL; L39209; AAC28750.1; -.
DR   HSSP; P00157; 1BE3.
DR   InterPro; IPR000179; Cyt_b.
DR   Pfam; PF00032; cytochrome_b_C; 1.
DR   Pfam; PF00033; cytochrome_b_N; 1.
DR   PROSITE; PS00190; CYTOCHROME_B_HEME; 1.
DR   PROSITE; PS00192; CYTOCHROME_B_HEME; 1.
DR   PROSITE; PS00193; CYTOCHROME_B_QO; 1.
KW   Mitochondrion; Electron transport; Respiratory chain; Transmembrane;
KW   Heme.
FT   METAL      83     83   IRON 1 (HEME B562 AXIAL LIGAND).
FT   METAL      97     97   IRON 2 (HEME B566 AXIAL LIGAND).
FT   METAL     182    182   IRON 2 (HEME B566 AXIAL LIGAND).
FT   METAL     196    196   IRON 1 (HEME B562 AXIAL LIGAND).
FT   CONFLICT  130    130   G -> S (IN REF. 2).
SQ   SEQUENCE   379 AA;   42606 MW;   6CAC4C6044EED898 CRC64;
```

```
Query Match       5.1%;   Score 98.5;   DB 1;   Length 379;
Best Local Similarity   20.4%;   Pred. No. 2.6;
Matches 69;  Conservative 42;  Mismatches 111;  Indels 117;  Gaps 17;

Qy   80  ICLLFMIGEVVGGYLA----------HSLAVWTDAAHLLTDF------------ASM-LISLF  119
             |||| : |:|| ||:|                 :||:                    |:| |:
Db   39  ICLILQI-LTGLFLAMHYTSDTTTAFSSVTHICRDVNYGWIIRYLHANGASMFFICLYM      96

Qy  120  ---SLWMSSRPATKTMNFG--------------WQRAEILGALVSLSIWVVTGV         157
               :|                            ||:|||||| :::
Db   97  HVGRGLYYGSYTFTETWNIGIILLFTVMATAFMGYVLPWGQMSFWGATV----ITNLLSA    152

Qy  158  LVILAVERL--ISGDYEIDGGTM-------LITSGCAVAVNIIMGLTLHQSGHGHSGTT    208
             |: |  || ||||:|:|||||      |||||||||||||||||||||||:|:|:||:
Db  153  IPYIGADLVEWIWGGFSVDKATLTRFFAFHFILPFVVSALATVHLLFLHETGSNNPSGIT    212

Qy  209  NQOEENPSVRAAFIHVIGDFMQSMG----VLVAAYILYPKPE----------YKYVDPICT--  255
             :::  |   |:|                ||
Db  213  SDSDKIP-------FHPYYIKDIGCALLLILVTLLVLFSPDLLGDPDNYTPANPLSTPP     266

Qy  256  ------FVFSIL--------VLGTTLTILRDVILVLME-GTPKGVDFTAVRDLILS        296
             |:|| :||||||| :| :|||||:                |
Db  267  HIKPEWYFLPAYAILRSIPNKLGGVLALVLSILIALMPLLHTSKQRGMMFRPISQCL--    324

Qy  297  VEGVEALHSLHIWALTV-----------AQPVLSVHIAIAQ                      326
             | : ||:|:| ||:||:           :||: ||:|: |
Db  325  ----------FWLLVADLLTLTWIGGQPVEHPVITIGQ                         352
```

RESULT 32

```
ID   CYB_PHATA    STANDARD;    PRT;    381 AA.
AC   Q35673;
DT   15-JUL-1998 (Rel. 36, Created)
DT   15-JUL-1998 (Rel. 36, Last sequence update)
DT   30-MAY-2000 (Rel. 39, Last annotation update)
DE   Cytochrome B.
GN   MTCYB OR COB OR CYTB.
OS   Phascogale tapoatafa (Common wambenger).
OG   Mitochondrion.
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Metatheria; Dasyuromorphia; Dasyuridae; Phascogale.
OX   NCBI_TaxID=9293;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=93096825; PubMed=1361058;
RA   Krajewski C., Driskell A.C., Baverstock P.R., Braun M.J.;
```

RT   "Phylogenetic relationships of the thylacine (Mammalia: Thylacinidae)
RT   among dasyuroid marsupials: evidence from cytochrome b DNA
RT   sequences.";
RL   Proc. R. Soc. Lond., B, Biol. Sci. 250:19-27(1992).
-!- FUNCTION: COMPONENT OF THE UBIQUINOL-CYTOCHROME C REDUCTASE
CC   COMPLEX (COMPLEX III OR CYTOCHROME B-C1 COMPLEX), WHICH IS A
CC   RESPIRATORY CHAIN THAT GENERATES AN ELECTROCHEMICAL POTENTIAL
CC   COUPLED TO ATP SYNTHESIS.
-!- COFACTOR: TWO HEME GROUPS (B562 AND B566) WHICH ARE NOT COVALENTLY
CC   BOUND TO THE PROTEIN.
-!- SUBUNIT: THE MAIN SUBUNITS OF COMPLEX B-C1 ARE: CYTOCHROME B,
CC   CYTOCHROME C1 AND THE RIESKE PROTEIN.
-!- SIMILARITY: BELONGS TO THE CYTOCHROME B FAMILY.

DR   EMBL; M99459; AAB40404.1; -.
DR   InterPro; IPR000179; Cyt_b_b6.
DR   Pfam; PF00032; cytochrome_b_C; 1.
DR   Pfam; PF00033; cytochrome_b_N; 1.
DR   PROSITE; PS00192; CYTOCHROME_B_HEME; 1.
DR   PROSITE; PS00193; CYTOCHROME_B_QO; 1.
KW   Heme.
KW   Mitochondrion; Electron transport; Respiratory chain; Transmembrane;
FT   METAL   83   83   IRON 1 (HEME B562 AXIAL LIGAND).
FT   METAL   97   97   IRON 1 (HEME B566 AXIAL LIGAND).
FT   METAL   182   182   IRON 2 (HEME B562 AXIAL LIGAND).
FT   METAL   196   196   IRON 2 (HEME B566 AXIAL LIGAND).
SQ   SEQUENCE   381 AA; 42755 MW; B8F15EBF4D5541C7 CRC64;

Query Match   5.1%;
Best Local Similarity   19.3%; Pred. No. 3.1;
Matches   67; Conservative   43; Mismatches 104; Indels 133; Gaps   16;
Score 97.5; DB 1; Length 381;

Qy   80 ICLLFMIGEVVGGYLA------HSLAVWTDAAHLLTDF-------------ASM-LISLF- 119
Db   39 ICLMIQI--LTGFFLAMHYTSDTLTAFTSVAHICRDVNYGWLLRNLHANGASMFFMCLFL 96

Qy 120 ----SLRMSSRPATKTMNFG-----------WQRAEILGALVSVSLIWVVTGV 157
Db   97 HVGLGIYYGSYLIKETWNIGVILLLTVWATAFVGYVLPWGQMSFWGAT------VITNL 149

Qy 158 L-----VYLAVERLISGDYEIDGGTM--------LITSGCAVAVNIIMGLTLHQSGHGHSH 205
Db 150 LSAIPYIGTTLAEWIWGGFAVDKATLTRFFAFHFILPFIIVALAIVHLLFHETGSNNPS 209

Qy 206 GTTNQEENPSVRAAFIHVIGDFMQSMGVLVAAYIL------------- 241
Db 210 GINPDSDKIP-----FHPYYIKDALGAVLLLVLLLLALFSPDSLGDPDNFSPANPLN 263

Qy 242 --YFKPEYKYVDPICTFVFSIL------VLGTTLTILRDVILVLM-----EGTPKGVDFT 288
Db 264 TPPHIKPEWYFL-------FAYAILRSIPNKLGGVLALLASILILLIPLLHTANQRSMFR 318

Qy 289 AVRDLLSVEGVEALHSLHIWALTV--------AQPVLSVHIAIAQ 326
Db 319 PVSQTL-----------FWILTADLITLTWIGGQPVEQPFIIGQ 352

RESULT 33
ID   ZNT4_BOVIN   STANDARD;   PRT;   50 AA.
AC   Q9TTF3;
DT   30-MAY-2000 (Rel. 39, Created)
DT   30-MAY-2000 (Rel. 39, Last sequence update)
DT   15-JUN-2002 (Rel. 41, Last annotation update)
DE   Zinc transporter 4 (ZnT-4) (Fragment).

GN   SLC30A4 OR ZNT4.
OS   Bos taurus (Bovine).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovoidea;
OC   Bovidae; Bovinae; Bos.
OX   NCBI_TaxID=9913;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=20078377; PubMed=10612253;
RA   Tammen I., Warren W.C., Raadsma H.W.;
RT   "Physical and linkage mapping of the bovine zinc transporter 4 (ZNT4)
RT   gene to chromosome 10.";
RL   Anim. Genet. 30:474-475(1999).
-!- FUNCTION: PROBABLY INVOLVED IN ZINC TRANSPORT OUT OF THE
CC   CYTOPLASM, MAY BE BY SEQUESTRATION INTO AN INTRACELLULAR
CC   COMPARTMENT (BY SIMILARITY).
-!- SUBCELLULAR LOCATION: Integral membrane protein.
CC   LOCALIZED IN ENDOSOMAL VESICLE MEMBRANE (PROBABLE).
-!- DOMAIN: CONTAINS A HISTIDINE-RICH REGION WHICH IS A LIGAND FOR
CC   ZINC, AND AN ASPARTIC ACID-RICH REGION WHICH IS A POTENTIAL LIGAND
CC   FOR ZINC (BY SIMILARITY).
-!- SIMILARITY: BELONGS TO THE CATION DIFFUSION FACILITATOR FAMILY OF
CC   TRANSPORTERS (CDF, TC 2.A.4). SLC30A SUBFAMILY.

DR   EMBL; AF103904; AAF18866.1; -.
DR   InterPro; IPR002524; Cation_efflux.
DR   Pfam; PF01545; Cation_efflux; 1.
KW   Zinc transport; Transport; Transmembrane; Multigene family.
FT   NON_TER   1   1
FT   DOMAIN   2   8   HIS-RICH.
FT   NON_TER   50   50
SQ   SEQUENCE   50 AA; 5159 MW; EAD57B8C21C65B60 CRC64;

Query Match   5.0%;
Best Local Similarity   44.7%; Pred. No. 0.37;
Matches   21; Conservative   8; Mismatches   6; Indels   12; Gaps   2;
Score 97; DB 1; Length 50;

Qy 201 HGHSHG-------TT-----NQQEENPSVRAAFIHVIGDFMQSMGVL 235
Db   4 HAHSHSLPSNSPTTGPRCGHNQGQDSLAVRAAFVHALGDLVQSVGVL 50

RESULT 34
ID   ARCD_PSEAE   STANDARD;   PRT;   482 AA.
AC   P18275;
DT   01-NOV-1990 (Rel. 16, Created)
DT   01-NOV-1990 (Rel. 16, Last sequence update)
DT   15-JUN-2002 (Rel. 41, Last annotation update)
DE   Arginine/ornithine antiporter.
GN   ARCD OR PA170.
OS   Pseudomonas aeruginosa.
OC   Bacteria; Proteobacteria; gamma subdivision; Pseudomonadaceae;
OC   Pseudomonas.
OX   NCBI_TaxID=287;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=90236296; PubMed=2158926;
RA   Luethi E., Baur H., Gamper M., Brunner F., Villeval D., Mercenier A.,
RA   Haas D.;
RT   "The arc operon for anaerobic arginine catabolism in Pseudomonas
RT   aeruginosa contains an additional gene, arcD, encoding a membrane
RT   protein.";
RL   Gene 87:37-43(1990).

```
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=ATCC 15692 / PAO1;
RX   MEDLINE=20437337; PubMed=10984043;
RA   Stover C.K., Pham X.-Q.T., Erwin A.L., Mizoguchi S.D., Warrener P.,
RA   Hickey M.J., Brinkman F.S.L., Hufnagle W.O., Kowalik D.J., Lagrou M.,
RA   Garber R.L., Goltry L., Tolentino E., Westbrock-Wadman S., Yuan Y.,
RA   Brody L.L., Coulter S.N., Folger K.R., Kas A., Larbig K., Lim R.M.,
RA   Smith K.A., Spencer D.H., Wong G.K.-S., Wu Z., Paulsen I.T.,
RA   Reizer J., Saier M.H., Hancock R.E.W., Lory S., Olson M.V.;
RT   "Complete genome sequence of Pseudomonas aeruginosa PAO1, an
RT   opportunistic pathogen.";
RL   Nature 406:959-964(2000).
RN   [3]
RP   CHARACTERIZATION.
RX   MEDLINE=92165731; PubMed=1311296;
RA   Verhoogt H.J., Smit H., Abee T., Gamper M., Driessen A.J., Haas D.,
RA   Konings W.N.;
RT   "arcD, the first gene of the arc operon for anaerobic arginine
RT   catabolism in Pseudomonas aeruginosa, encodes an arginine-ornithine
RT   exchanger.";
RL   J. Bacteriol. 174:1568-1573(1992).
-!-  FUNCTION: CATALYZES AN ELECTRONEUTRAL EXCHANGE BETWEEN ARGININE
CC   AND ORNITHINE TO ALLOW HIGH-EFFICIENCY ENERGY CONVERSION IN THE
CC   ARGININE DEIMINASE PATHWAY.
-!-  SUBCELLULAR LOCATION: Integral membrane protein. Inner membrane
CC   (Potential).
-!-  SIMILARITY: BELONGS TO THE ARCD/CADB/GERAB/LYSI FAMILY OF
CC   PERMEASES.
```

```
DR   EMBL; M33223; AAA25719.1; -.
DR   EMBL; AE004930; AAG05555.1; -.
DR   PIR; JH0110; JH0110.
DR   InterPro; IPR002293; AA/rel_permease.
DR   InterPro; IPR004754; R/O_antiport.
DR   InterPro; IPR004841; Permease.
Pfam; PF00324; aa_permeases; 1.
TIGRFAMs; TIGR00905; 2A0302; 1.
KW   Transport; Antiport; Amino-acid transport; Transmembrane;
KW   Inner membrane; Complete proteome.
FT   TRANSMEM      11     30       POTENTIAL.
FT   TRANSMEM      40     61       POTENTIAL.
FT   TRANSMEM      94    118       POTENTIAL.
FT   TRANSMEM     126    145       POTENTIAL.
FT   TRANSMEM     163    184       POTENTIAL.
FT   TRANSMEM     204    221       POTENTIAL.
FT   TRANSMEM     236    258       POTENTIAL.
FT   TRANSMEM     281    305       POTENTIAL.
FT   TRANSMEM     335    352       POTENTIAL.
FT   TRANSMEM     365    389       POTENTIAL.
FT   TRANSMEM     405    421       POTENTIAL.
FT   TRANSMEM     428    444       POTENTIAL.
FT   TRANSMEM     459    475       POTENTIAL.
SQ   SEQUENCE 482 AA; 52026 MW; C7FCIA73233FB0607 CRC64;

Query Match 5.0%; Score 97; DB 1; Length 482;
Best Local Similarity 20.0%; Pred. No. 4.3;
Matches 62; Conservative 50; Mismatches 86; Indels 112; Gaps 12;

Qy   65 PKKGKAQRQLVASAICLLEMIGEVV--GGYLAHSLAVMTDAAHLITDFASMLISLFS-- 120
         | |   || |:  | |::  :|     ||   ::| |  ||  || ||  :
Db  119 PIFGKGDTVAAIVCASVLLWALHFLVLRGIKEBAAFINTVTTVAKVVPLFLFILCLFAFK 178

Qy  121 ------LWMSSRPAFKTMNFGWQRAEILGALVSVLSIWVVTGVLYVLYLAVERLISGDYEI 173
         :| |                ||   :  | |  :: :| :::  :|  | |
Db  179 LDIFTADIWGKSNPD------------LGSVMNQVRNMLVTWVVFIGIE------ 216

Qy  174 DGGTMLITS-------GCAAVNIIMGLTLHQSGHGHSHGTINQQE----- ENPSVRAAF 221
          |: :| | | |::   | ||::| |  ||  :|| |: :||:|  ::||:|
Db  217 --GASIFSSRAEKRSDVGKRATVIGFITVLLLVLVNVLSMGVMVTQPELAKLIQNPSMALVL 274

Qy  222 IHVIGDF--MQSMGVLVA-------AYIL------ YFKPEY------ 247
         ||:|: | :|:   :|| ::| | |:
Db  275 EHVVGHWGAVLISVGLLISLLGALLSWVLLCAEIMFAAAKDHTMPEFLRRENANQVPANA 334

Qy  248 KYVDPICTFEVFSILVLGTLTILRDVILVLMEGTPKGVD------------FTA 289
         || ||  |::  |  |:  || :|:: :|:|:|:|| |    :|
Db  335 LWLTNICVQVFLVVFFTS----------GDPDGMDPYTKMLLLATSMLLIPYFWSA 381

Qy  290 VRDLLLSVEG 299
         ||:|:|| :|
Db  382 AYGLLLTLKG 391
```

RESULT 35
TSCC_HUMAN

```
ID   TSCC_HUMAN     STANDARD;      PRT;   1021 AA.
AC   P55017;
DT   01-OCT-1996 (Rel. 34, Created)
DT   01-OCT-1996 (Rel. 34, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   Thiazide-sensitive sodium-chloride cotransporter (NA-CL symporter).
GN   SLC12A3 OR TSC.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A. AND VARIANTS GS.
RX   MEDLINE=96122035; PubMed=8528245;
RA   Simon D.B., Nelson-Williams C., Bia M.J., Ellison D., Karet F.E.,
RA   Molina A.M., Vaara I., Iwata F., Cushner H.M., Koolen M., Gainza F.J.,
RA   Gitelman H.J., Lifton R.P.;
RT   "Gitelman's variant of Bartter's syndrome, inherited hypokalaemic
RT   alkalosis, is caused by mutations in the thiazide-sensitive Na-Cl
RT   cotransporter.";
RL   Nat. Genet. 12:24-30(1996).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Kidney;
RX   MEDLINE=97001149; PubMed=8812482;
RA   Mastroianni N., de Fusco M., Zollo M., Arrigo G., Zuffardi O.,
RA   Bettinelli A., Ballabio A., Casari G.;
RT   "Molecular cloning, expression pattern, and chromosomal localization
RT   of the human Na-Cl thiazide-sensitive cotransporter (SLC12A3).";
RL   Genomics 35:486-493(1996).
-!-  FUNCTION: ELECTRICALLY SILENT TRANSPORTER SYSTEM WHICH IS A
CC   MEDIATOR OF SODIUM AND CHLORIDE REABSORPTION.
-!-  SUBCELLULAR LOCATION: Integral membrane protein.
-!-  TISSUE SPECIFICITY: PREDOMINANT IN KIDNEY.
-!-  DISEASE: DEFECTS IN SLC12A3 ARE THE CAUSE OF GITELMAN'S SYNDROME
CC   (GS), AN AUTOSOMAL RECESSIVE DISEASE CHARACTERIZED BY DIVERSE
CC   ABNORMALITIES IN ELECTROLYTE HOMEOSTASIS INCLUDING HYPOKALAEMIC
CC   METABOLIC ALKALOSIS. GS IS A SUBSET OF BARTTER'S SYNDROME.
-!-  SIMILARITY: BELONGS TO THE SLC12A FAMILY OF TRANSPORTERS.
```

```
DR   EMBL; U44128; AAC50355.1; -.
DR   EMBL; X91220; CAA62613.1; -.
DR   Genew; HGNC:10912; SLC12A3.
DR   MIM; 600968; -.
```

```
DR   MIM; 263800; -.
DR   InterPro; IPR002293; AA/rel_prmease1.
DR   InterPro; IPR004842; KC1_cotransport.
DR   InterPro; IPR002948; NaCl_transporter.
DR   PRINTS; PR01230; NACLTRNSPORT.
DR   TIGRFAMs; TIGR00930; 2a30; 1.
KW   Transport; Transmembrane; Glycoprotein; Disease mutation.
FT   DOMAIN        1    135  CYTOPLASMIC (POTENTIAL).
FT   TRANSMEM    136    156  POTENTIAL.
FT   TRANSMEM    157    158  POTENTIAL.
FT   DOMAIN      159    179  EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM    180    218  POTENTIAL.
FT   DOMAIN      219    239  CYTOPLASMIC (POTENTIAL).
FT   TRANSMEM    240    261  POTENTIAL.
FT   DOMAIN      262    282  EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM    283    307  POTENTIAL.
FT   DOMAIN      308    339  CYTOPLASMIC (POTENTIAL).
FT   TRANSMEM    340    360  POTENTIAL.
FT   DOMAIN      361    377  EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM    378    398  POTENTIAL.
FT   DOMAIN      399    452  EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM    453    473  POTENTIAL.
FT   DOMAIN      474    511  EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM    512    532  POTENTIAL.
FT   DOMAIN      533    534  CYTOPLASMIC (POTENTIAL).
FT   TRANSMEM    535    555  POTENTIAL.
FT   DOMAIN      556    577  EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM    578    598  POTENTIAL.
FT   DOMAIN      599    660  CYTOPLASMIC (POTENTIAL).
FT   TRANSMEM    661    681  POTENTIAL.
FT   DOMAIN      682   1021  CYTOPLASMIC (POTENTIAL).
FT   CARBOHYD    406    406  N-LINKED (GLCNAC...) (POTENTIAL).
FT   CARBOHYD    426    426  N-LINKED (GLCNAC...) (POTENTIAL).
FT   VARIANT     209    209  R -> W (IN GS).
FT                           /FTId=VAR_007113.
FT   VARIANT     349    349  P -> L (IN GS).
FT                           /FTId=VAR_007114.
FT   VARIANT     421    421  C -> R (IN GS).
FT                           /FTId=VAR_007115.
FT   VARIANT     486    486  D -> N (IN GS).
FT                           /FTId=VAR_007116.
FT   VARIANT     496    496  G -> C (IN GS).
FT                           /FTId=VAR_007117.
FT   VARIANT     561    561  MISSING (IN GS).
FT                           /FTId=VAR_007118.
FT   VARIANT     588    588  A -> V (IN GS).
FT                           /FTId=VAR_007119.
FT   VARIANT     630    630  G -> V (IN GS).
FT                           /FTId=VAR_007120.
FT   VARIANT     655    655  R -> H (IN GS).
FT                           /FTId=VAR_007121.
FT   VARIANT     655    655  R -> L (IN GS).
FT                           /FTId=VAR_007122.
FT   VARIANT     728    728  A -> T (IN GS).
FT                           /FTId=VAR_007123.
FT   VARIANT     741    741  G -> R (IN GS).
FT                           /FTId=VAR_007124.
FT   VARIANT     850    850  L -> P (IN GS).
FT                           /FTId=VAR_007125.
FT   VARIANT     955    955  R -> Q (IN GS).
FT                           /FTId=VAR_007126.
FT   CONFLICT    459    460  AG -> VV (IN REF. 2).
FT   CONFLICT    766    766  E -> D (IN REF. 2).
FT   CONFLICT    807    807  V -> GARPSVSGAL (IN REF. 1).
SQ   SEQUENCE   1021 AA;  113138 MW;  D7ECE53DA6233821 CRC64.
```

```
Qy   57 KGPDSHCDPKKGKAQRQLYVASAICLLFMIGEVVGGYLAHSLAVMTDAAHLLTDFASMLI 116
         ||  |:||   |  || |:  |  |:||||   |:|:|:   | ||||  :|:||||  :||:|
```

```
Query Match          5.0%;
Best Local Similarity  19.5%;
Matches   42;  Conservative   51;  Mismatches   59;  Indels   63;  Gaps   8;
Score 96.5;   Pred. No. 11;   DB 1;   Length 1021;
```

```
Db   497 KGYGKNKEPVRG--------YLLAYAIAVAFII--------IAELNTIAPIISNF--FLC 538
             |:|  |:                                   : :                     || :
Qy   117 SLFSLWMSSRPATIKTMNFGWQR--------AEILGALVSVLSIWVVT--------GVLVY 160

Qy   497 SYALINFSCFHASITNSPGWRPSFQYYNKWAALFGAIISVVIMFLLTWWAALIAIGVVLF 598
             |:||   | :||   | ||    :||:| | ||:|  :|:||||   |::||:|
Db   539 ...

Qy   161 LAVERLISGDYEIDGGTMLITSGCAVAVNIIMGLLHQSGHGHSHGTTNQQEEN------- 214
             |::|         :||  |:|||         || |:|  |:||::||
Db   599 LLL-TVIYKPEVNNGSSVQAGSYNLALSYSVGL--------------NEVEDHIKNYRP 643

Qy   215 --------PSVRAAFTHVIGDFMQSMGVLVAAYIL 241
                         |:||   :||  |||| |::||:|
Db   644 QCLVLTGPPNFRPALVDFGVGTFTRNLSLMICGHVL 678
```

```
RESULT 36
CYB_HYLLA

ID   CYB_HYLLA    STANDARD;    PRT;    380 AA.
AC   Q95711; Q34733;
DT   01-NOV-1997 (Rel. 35, Created)
DT   01-NOV-1997 (Rel. 35, Last sequence update)
DT   15-JUN-2002 (Rel. 41, Last annotation update)
DE   Cytochrome B.
GN   MTCYB OR COB OR CYTB.
OS   Hylobates lar (Common gibbon).
OG   Mitochondrion.
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hylobatidae; Hylobates.
OX   NCBI_TaxID=9580;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=Ester;
RA   Garza J.C., Woodruff D.S.;
RT   "A phylogenetic study of the gibbons (Hylobates) using DNA obtained
RT   noninvasively from hair.";
RL   Mol. Phylogenet. Evol. 1:202-210(1992).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=Hair;
RX   MEDLINE=94115700; PubMed=1342936;
RA   Arnason U., Gullberg A., Xu X.;
RT   "A complete mitochondrial DNA molecule of the white-handed gibbon,
RT   Hylobates lar, and comparison among individual mitochondrial genes of
RT   all hominoid genera.";
RL   Hereditas 124:185-189(1996).
CC   -!- FUNCTION: COMPONENT OF THE UBIQUINOL-CYTOCHROME C REDUCTASE
CC       COMPLEX (COMPLEX III OR CYTOCHROME B-C1 COMPLEX), WHICH IS A
CC       RESPIRATORY CHAIN THAT GENERATES AN ELECTROCHEMICAL POTENTIAL
CC       COUPLED TO ATP SYNTHESIS.
CC   -!- COFACTOR: TWO HEME GROUPS (B562 AND B566) WHICH ARE NOT COVALENTLY
CC       BOUND TO THE PROTEIN.
CC   -!- SUBUNIT: THE MAIN SUBUNITS OF COMPLEX B-C1 ARE: CYTOCHROME B,
CC       CYTOCHROME C1 AND THE RIESKE PROTEIN.
CC   -!- SIMILARITY: BELONGS TO THE CYTOCHROME B FAMILY.
```

```
DR   EMBL; X99256; CAA67640.1; -.
DR   EMBL; L02772; AAA31789.1; -.
DR   InterPro; IPR000179; Cyt_b-b6.
DR   Pfam; PF00032; cytochrome_b_C; 1.
DR   Pfam; PF00033; cytochrome_b_N; 1.
DR   PROSITE; PS00190; CYTOCHROME_B_HEME; 1.
DR   PROSITE; PS00192; CYTOCHROME_B_N; 1.
DR   PROSITE; PS00193; CYTOCHROME_B_QO; 1.
KW   Electron transport; Mitochondrion; Respiratory chain; Transmembrane;
KW   Heme.
FT   METAL        83     83  IRON 1 (HEME B562 AXIAL LIGAND).
```

```
FT   METAL        97     97       IRON 2 (HEME B566 AXIAL LIGAND).
FT   METAL       182    182       IRON 2 (HEME B562 AXIAL LIGAND).
FT   METAL       196    196       IRON 1 (HEME B566 AXIAL LIGAND).
SQ   SEQUENCE   380 AA;  42619 MW;  254DD2469287587 CRC64;

Query Match      5.0%;  Score 96;  DB 1;  Length 380;
Best Local Similarity   20.2%;  Pred. No. 4;
Matches   76;  Conservative 51;  Mismatches 96;  Indels 154;  Gaps 21;

Qy   81  CLLFMIGEVVGGYLA-----HSLAVWTDAAHLLTDF--------------------ASM-LISLF-- 119
         |||: |  |: | |||                                       ||| ||| |
Db   40  CLILQI--ITGLFLAMHYTDPASTAFSSVAHITRDVNYGWIIRYLHANGASMFFICLFLH 97

Qy  120  ---SLMMSSRPATKTMNFG------------------WQRAEILGALVSVLSIWVVTGVL 158
                                             ||:  |: | |:
Db   98  IGRGLYGSFLYLETWNIGIILLLATMATAFPMGCYVLPWGQMSFWGATV-- 153

Qy  159  VYLAVE--RLISGDYEIDGGTM----------LITSGCAVAVNIIMGLTLHQSGHGHS 204
         : |    : |  |:                   ||| :   |  |: ||
Db  154  PYIGTDLVQWVWGGYSVDNATLTRFFTFHFILPFIITALAALHL-----LFLHETGSNNP 208

Qy  205  HGTTNQQEENPSVRAAFIH--------VIGDFMQSMGVLVAAYILYFKPE------YKYV 250
         |: |         :|| |        :||::|:  :|:|
Db  209  LGISSQPD------KIAF-HPYYTIKDILGFLL---LLMLMSLVLFSPDLLGDPSNYTQA 259

Qy  251  DPICT--------------FVFSIL-----VLGTTLTLIRDVIVLVIMEGTPKGVDFTAVRDLL 294
         :||:                |||:     |||::||:|||||:|:|||||:|||||||||||
Db  260  NPLNTPPHIRPEWYFLFAYAILRSVPNKLGGVLALLLSILILAM-------------- 303

Qy  295  LSVEGVEALHS-----------------LHIWALTV------AQPVLSVHIAIAQNTDAQ- 331
         |:|                        ||||||:       ||||||||:||||||:||
Db  304  -----IPALHTAKQQSMMFRPLSQLTVWLLVWNLLILTWIGGQPVSVPFIITGQVASALY 358

Qy  332  ----AVLKTASSRLQGK 344
             ||||||||||||
Db  359  FTTIILVLMPAASLIENK 375
```

RESULT 37
MNTH_SALTI

```
ID   MNTH_SALTI    STANDARD;     PRT;    413 AA.
AC   Q824X5;
DT   15-JUN-2002 (Rel. 41, Created)
DT   15-JUN-2002 (Rel. 41, Last sequence update)
DT   15-JUN-2002 (Rel. 41, Last annotation update)
DE   Manganese transport protein mntH.
GN   MNTH OR STY2649.
OS   Salmonella typhi.
OC   Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC   Salmonella.
OX   NCBI_TaxID=601;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=CT18;
RX   MEDLINE=21534947; PubMed=11677608;
RA   Parkhill J., Dougan G., James K.D., Thomson N.R., Pickard D., Wain J.,
RA   Churcher C., Mungall K.L., Bentley S.D., Holden M.T.G., Sebaihia M.,
RA   Baker S., Basham D., Brooks K., Chillingworth T., Connerton P.,
RA   Cronin A., Davis P., Davies R.M., Dowd L., White N., Farrar J.,
RA   Feltwell T., Hamlin N., Haque A., Hien T.T., Holroyd S., Jagels K.,
RA   Krogh A., Larsen T.S., Leather S., Moule S., O'Gaora P., Parry C.,
RA   Quail M., Rutherford K., Simmonds M., Skelton J., Stevens K.,
RA   Whitehead S., Barrell B.G.;
RT   "Complete genome sequence of a multiple drug resistant Salmonella
RT   enterica serovar Typhi CT18.";
RL   Nature 413:848-852(2001).
CC   -!- FUNCTION: H(+)-stimulated, highly selective, manganese uptake
CC       system. Can also transport cadmium, cobalt, iron, zinc and to a
CC       lesser extent nickel and copper. Involved in response to reactive
CC       oxygen (By similarity).
CC   -!- SUBCELLULAR LOCATION: Integral membrane protein. Inner membrane
CC       (Probable).
CC   -!- SIMILARITY: BELONGS TO THE NRAMP FAMILY.
```

```
DR   EMBL; AL627274; CAD07646.1; -.
DR   TIGRFAMs; TIGR01197; nramp; 1.
KW   Complete proteome.
KW   Transport; Symport; Manganese; Transmembrane; Inner membrane;
FT   TRANSMEM     19     39       POTENTIAL.
FT   TRANSMEM     48     68       POTENTIAL.
FT   TRANSMEM     94    114       POTENTIAL.
FT   TRANSMEM    122    142       POTENTIAL.
FT   TRANSMEM    156    176       POTENTIAL.
FT   TRANSMEM    196    216       POTENTIAL.
FT   TRANSMEM    241    261       POTENTIAL.
FT   TRANSMEM    290    310       POTENTIAL.
FT   TRANSMEM    329    349       POTENTIAL.
FT   TRANSMEM    350    370       POTENTIAL.
FT   TRANSMEM    389    409       POTENTIAL.
SQ   SEQUENCE   413 AA;  44414 MW;  B8CEC003C0F9F3C3 CRC64;

Query Match      5.0%;  Score 96;  DB 1;  Length 413;
Best Local Similarity   19.4%;  Pred. No. 4.3;
Matches   79;  Conservative 56;  Mismatches 119;  Indels 154;  Gaps 17;

Qy   60  DSHCDPKGKAQRQLYVA-SAICLLFMIGEVVGGYLAHSLAVWTDAAHLL-------TDFA 112
         :                  :|  ||||||||||||| |||||||||||       ||||
Db    3  DNRVENSSGRAARKLRLALMGPAFIAAIGYIDPGNFATNIQAGASFGYQLLWVVWANLM 62

Qy  113  SMLISLFSLWMS----------------SRPATKTMNFGWQRAEI-----------LGAL-- 145
         :|||                        |||||||||||||||            ||||
Db   63  AMLIQILSAKLGIATGKNLAEQIRDHYPRP---VVWFVWVQAEIIAMATDLAEFIGAAIG 119

Qy  146  -----VSVLSIWVTGVLVLYLAVERLISGDYEID---GGTMLITSGGAV--------- 186
              |:||:|||:||||||||||||||:|||||||   |||:||||||||
Db  120  FKLILGVSLLQGAVLTGIATFLILMLQRRGQKPLEKVIGGLLLFVAAAYIVELFFSQPDM 179

Qy  187  -----------AVNI---IMGLT*------LHQSGHGHSHGTTNQQEENPSV-RA 219
                     |:|:   |||:       |||||||||||||||||||||| ||
Db  180  AQLGKGMVIPALNPEAVFLAAGVLGATIMPHVIYLHSSLTQHLHGGTRQQRYSATKWDV 239

Qy  220  AFIHVIGDFMQSMGVLVAAYILYFK-------PEYKYVDPICTFV----------- 257
         |:|::||||||||||||||||||||       |||||||||||||
Db  240  AIAMTIAGFVNLAMMATAAAAAFHRSGHTGIADLDQAYLTLEPLLSHAAATVFGLSLVAAG 299

Qy  258  FSILVLGT------------------TLTLIRDVILVLMEGTPKGVDFTAVRDLLLS 296
         |||:|||:                  ||||||||:|||||||||||||||||||||
Db  300  LSSTVVGTLAGQVVMQGFVRFHIPLWVRRTITIMLPSFIVILM-----GLDPTRI------ 348

Qy  297  VEGVEALHSLHIWALTVAQPVLSVHIAIAQNTDAQAVLKTASSRLQGK 344
         :  |  | |:|:  :   ||||||||:||||:| |||||||||||||
Db  349  -----------LVMSQVLLSFGIALAL---VPLLIFTSNATLMGE 379
```

RESULT 38
MNTH_SALTY

```
ID   MNTH_SALTY    STANDARD;     PRT;    413 AA.
AC   Q9RPF4;
DT   16-OCT-2001 (Rel. 40, Created)
DT   16-OCT-2001 (Rel. 40, Last sequence update)
DT   15-JUN-2002 (Rel. 41, Last annotation update)
DE   Manganese transport protein mntH.
GN   MNTH OR STM2408.
OS   Salmonella typhimurium.
OC   Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC   Salmonella.
OX   NCBI_TaxID=602;
RN   [1]
```

RP   SEQUENCE FROM N.A., AND CHARACTERIZATION.
RC   STRAIN=LT2 / MM1255;
RX   MEDLINE=20305051; PubMed=10844693;
RA   Kehres D.G., Zaharik M.L., Finlay B.B., Maguire M.E.;
RT   "The NRAMP proteins of Salmonella typhimurium and Escherichia coli are
RT   selective manganese transporters involved in the response to reactive
RT   oxygen.";
RL   Mol. Microbiol. 36:1085-1100(2000).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=LT2 / SGSC1412 / ATCC 700720;
RX   MEDLINE=21534948; PubMed=11677609;
RA   McClelland M., Sanderson K.E., Spieth J., Clifton S.W., Latreille P.,
RA   Courtney L., Porwollik S., Ali J., Dante M., Du F., Hou S., Layman D.,
RA   Leonard S., Nguyen C., Scott K., Holmes A., Grewal N., Mulvaney E.,
RA   Ryan E., Sun H., Florea L., Miller W., Stoneking T., Nhan M.,
RA   Waterston R., Wilson R.K.;
RT   "Complete genome sequence of Salmonella enterica serovar Typhimurium
RT   LT2.";
RL   Nature 413:852-856(2001).
CC   -!- FUNCTION: H(+)-STIMULATED, HIGHLY SELECTIVE, MANGANESE UPTAKE
CC   SYSTEM. CAN ALSO TRANSPORT CADMIUM, COBALT, IRON, ZINC AND TO A
CC   LESSER EXTENT NICKEL AND COPPER. INVOLVED IN RESPONSE TO REACTIVE
CC   OXYGEN.
CC   -!- SUBCELLULAR LOCATION: Integral membrane protein. Inner membrane.
CC   -!- SIMILARITY: BELONGS TO THE NRAMP FAMILY.
CC   --
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   --
DR   EMBL; AF161317; AAD46616.1; -.
DR   EMBL; AE008808; AAL21308.1; -.
DR   StyGene; SG10755; mntH.
DR   InterPro; IPR001046; Nramp.
Pfam; PF01566; Nramp; 1.
DR   PRINTS; PR00447; NATRESASSCMP.
ProDom; PD001861; Nramp; 1.
DR   TIGRFAMs; TIGR01197; nramp; 1.
KW   Transport; Symport; Manganese; Transmembrane; Inner membrane;
KW   Complete proteome.
FT   TRANSMEM   19   39   POTENTIAL.
FT   TRANSMEM   48   68   POTENTIAL.
FT   TRANSMEM   94   114   POTENTIAL.
FT   TRANSMEM   122   142   POTENTIAL.
FT   TRANSMEM   156   176   POTENTIAL.
FT   TRANSMEM   196   216   POTENTIAL.
FT   TRANSMEM   241   261   POTENTIAL.
FT   TRANSMEM   290   310   POTENTIAL.
FT   TRANSMEM   329   349   POTENTIAL.
FT   TRANSMEM   350   370   POTENTIAL.
FT   TRANSMEM   392   412   POTENTIAL.
SQ   SEQUENCE   413 AA;   44400 MW; 953F2F03C0F9F3C3 CRC64;

Query Match   5.08%; Score 96; DB 1; Length 413;
Best Local Similarity   19.4%;
Matches 79; Conservative 56; Mismatches 119; Indels 154; Gaps 17;

Qy   60 DSHCDPKKGKAQRQLYVA-SAICLLFMIGEVVGGYLAHSLAVMTDAAHLL------TDFA 112
                |:|| | | ||: | | |:|||: | |||:|:| | | ||
Db   3 DNRVENSSGRAARKLRLALMGPAFIAAIGYIDPGNFATNIQAGASFGYQLLWVVWANLM 62

Qy   113 SMLISLFSLWMS-----------SRPATKTMNFGWQRAEI----------LGAL-- 145
            :||:||| |: ||: || | ||
Db   63 AMLIQILSAKLGIATGKNLAEQIRDHYPRP--VVWFWVQAEIIAMATDLAEFIGAAIG 119

Qy   146 -----VSVLSIWWTGVLVYLAVERLISGDYEID---GGTMLITSGCAV-------- 186
                 |:|: |:||: | | ||:|:| ||:||||
Db   120 FKLILGVSLLQGAVLTGIATFLIMLQRRGQRPLEKVIGGLLLFVAAAYIVELFFSQPDM 179

RESULT 39
YN8N_YEAST
ID   YN8N_YEAST   STANDARD;   PRT;   605 AA.
AC   P53735;
DT   01-OCT-1996 (Rel. 34, Created)
DT   01-OCT-1996 (Rel. 34, Last sequence update)
DT   15-JUL-1998 (Rel. 36, Last annotation update)
DE   Hypothetical 67.5 kDa protein in DBP6-COO2 intergenic region.
GN   YNR039C OR N3403.
OS   Saccharomyces cerevisiae (Baker's yeast).
OC   Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC   Saccharomycetales; Saccharomycetaceae; Saccharomyces.
OX   NCBI_TaxID=4932;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Pohl T.M.;
RL   Submitted (MAY-1996) to the EMBL/GenBank/DDBJ databases.
CC   -!- SUBCELLULAR LOCATION: Integral membrane protein (Potential).
CC   --
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   --
DR   EMBL; Z71654; CAA96319.1; -.
DR   SGD; S0005322; YNR039C.
KW   Hypothetical protein; Transmembrane.
FT   TRANSMEM   226   246   POTENTIAL.
FT   TRANSMEM   255   275   POTENTIAL.
FT   TRANSMEM   288   308   POTENTIAL.
FT   TRANSMEM   310   330   POTENTIAL.
FT   TRANSMEM   364   384   POTENTIAL.
FT   TRANSMEM   400   420   POTENTIAL.
FT   TRANSMEM   423   443   POTENTIAL.
FT   TRANSMEM   546   566   POTENTIAL.
SQ   SEQUENCE   605 AA;   67460 MW; 299A89B0E4FCDBC8 CRC64;

Query Match   5.08%; Score 96; DB 1; Length 605;
Best Local Similarity   20.7%;
Matches 52; Conservative 37; Mismatches 102; Indels 60; Gaps 9;

Qy   69 KAQRQLYVAS----AICLLFMIGEVVGGYLAHSLAVMTDAAHLLT-DFASMLI----- 116
            ||:| ||:|: | ||| |||: | |||:|:||||||| | || |||
Db   69 KAYIQLSIALQIFACLITFQVGHL-----YSWSNFITLSHFITYDIIGSLVIIFVENL 278

Qy   117 SLFSLWMSSRPATKIMNFGWQRAEILGALVSVLSIWWVTGVLVYLAVERLISGDYEIDGG 176
            ||:|| |:||||: | ||| || || :| | ||: | | ||:| |:
Db   279 SQFQVWFT--GTIIFPFGLNRIDVLLSFALAVSLCFVGLDLLFHIIEEFI---------- 326

Qy   177 TMLITSGCAVAVNIIMGLTLHQSGHGHGHSHGTTNQOEENPSV--------------RA 219
            ::||:||||| :| ||| ||:|:|:| :| | | |||: ||
Db   327 VLFVESGGSSLTNNHDHDEINEQIPHSHIANANDSQNENIILWYSILMINLVLSTLSLYKT 386

```
Qy    220  AFIHVIGDFMQSMGVLVAAYILYFKPEYKY------------------VDPICTFVFSILVLGTTLTI  269
                |||:|                                        ||:| |||   | |:| ||||
Db    387  FYANKXSNLKTKNPITITYAYL---FIYPLLLDLLSSISDYLATLVISSLILWHGLTI  443

Qy    270  LRDVILVLMEG  280
                ||:|  |
Db    444  ARWTSTVLLMG  454
```

RESULT 40

```
ID    NU5C_SYNP2    STANDARD;    PRT;    664 AA.
AC    P31971;
DT    01-JUL-1993 (Rel. 26, Created)
DT    01-JUL-1993 (Rel. 26, Last sequence update)
DT    01-NOV-1997 (Rel. 35, Last annotation update)
DE    NADH-plastoquinone oxidoreductase chain 5 (EC 1.6.5.3).
GN    NDHF.
OS    Synechococcus sp. (strain PCC 7002) (Agmenellum quadruplicatum).
OC    Bacteria; Cyanobacteria; Chroococcales; Synechococcus.
OX    NCBI_TaxID=32049;
RN    [1]
RP    SEQUENCE FROM N.A.
RX    MEDLINE=92207922; PubMed=1554697;
RA    Schluchter W.M., Bryant D.A.;
RT    "Molecular characterization of ferredoxin-NADP+ oxidoreductase in
RT    cyanobacteria: cloning and sequence of the petH gene of Synechococcus
RT    sp. PCC 7002 and studies on the gene product.";
RL    Biochemistry 31:3092-3102(1992).
RN    [2]
RP    SEQUENCE OF 1-20 FROM N.A.
RX    MEDLINE=93273700; PubMed=8501038;
RA    Schluchter W.M., Zhao J., Bryant D.A.;
RT    "Isolation and characterization of the ndhF gene of Synechococcus sp.
RT    strain PCC 7002 and initial characterization of an interposon
RT    mutant.";
RL    J. Bacteriol. 175:3343-3352(1993).
CC    -!- CATALYTIC ACTIVITY: NADH + plastoquinone = NAD(+) + plastoquinol.
CC    -!- SUBCELLULAR LOCATION: Integral membrane protein.
CC    -!- SIMILARITY: TO POLYPEPTIDE 5 OF THE NADH-UBIQUINOL OXIDOREDUCTASE
CC        OF CHLOROPLASTS OR MITOCHONDRIA.
CC    --------------------------------------------------------------------------
CC    This SWISS-PROT entry is copyright. It is produced through a collaboration
CC    between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC    the European Bioinformatics Institute. There are no restrictions on its
CC    use by non-profit institutions as long as its content is in no way
CC    modified and this statement is not removed. Usage by and for commercial
CC    entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC    or send an email to license@isb-sib.ch).
CC    --------------------------------------------------------------------------
DR    EMBL; M99378; AAA27311.1; -.
DR    EMBL; M86234; -; NOT_ANNOTATED_CDS.
DR    PIR; A40588; A40588.
DR    InterPro; IPR003916; NADHub_oxred5.
DR    InterPro; IPR001750; Oxidored_q1.
DR    InterPro; IPR001516; Oxidored_q1_N.
DR    Pfam; PF00361; oxidored_q1; 1.
DR    Pfam; PF00662; oxidored_q1_N; 1.
DR    PRINTS; PR01434; NADHDHGNASE5.
KW    Oxidoreductase; NAD; Plastoquinone; Transmembrane.
FT    TRANSMEM      7     27       POTENTIAL.
FT    TRANSMEM     39     59       POTENTIAL.
FT    TRANSMEM     91    111       POTENTIAL.
FT    TRANSMEM    120    140       POTENTIAL.
FT    TRANSMEM    144    164       POTENTIAL.
FT    TRANSMEM    187    207       POTENTIAL.
FT    TRANSMEM    219    239       POTENTIAL.
FT    TRANSMEM    258    278       POTENTIAL.
FT    TRANSMEM    290    310       POTENTIAL.
FT    TRANSMEM    327    347       POTENTIAL.
FT    TRANSMEM    352    372       POTENTIAL.
FT    TRANSMEM    395    415       POTENTIAL.
FT    TRANSMEM    420    440       POTENTIAL.
FT    TRANSMEM    495    515       POTENTIAL.
FT    TRANSMEM    541    561       POTENTIAL.
FT    TRANSMEM    643    663       POTENTIAL.
SQ    SEQUENCE   664 AA;  72926 MW;  ED60ACF947239BC5 CRC64;
```

Query Match   5.0%; Score 95.5; DB 1; Length 664;
Best Local Similarity   21.7%; Pred. No. 7.9;
Matches   54; Conservative   34; Mismatches 110; Indels 51; Gaps   8;

```
Qy     76  VASAICLLFMIGEVVGGYIAHSLAVMTDAAHLLTDFASMLISLFS-----------------  120
                : : :||| | :| :||:|           |      ||:|| :||||| : 
Db     99  IVTTVALLVMI--YTDGYMAHDPGYVRFYAYLSIFSSSMLGLVFSPNLVQVYIFWELVGM  156

Qy    121  ------LWMSSRPATKTMN--FGWQRAEILGALVSVLSIWVVTGVLVYLAVERLISGDY  171
                      | |     || | |  || |:  || |:: |  ||
Db    157  CSYLLIGFWYDRKAAADACQKAFVTNRVGDFGLLLGMLGLYWATG-----SFEFDLMGDR  211

Qy    172  EIDGGTMLLTSG-------CAVAVNIIMGLTLHQSGHGHSHGTTNQQEENPSVRAAFIHV  224
                :|          | :||:  :  | |  | :||  ||          ||
Db    212  LMD---LVSTGQISSLLAIVFAVLVFLG-PVAKSAQPFLHVWLPDAMEGPTPISALIHA  266

Qy    225  IGDEMQSMGVLVAAYILYFKPEYKYVDPICTFVFSILVLGTTLTILRDVILVLMEGTPKG  284
                |||     ||| : | : |  | |||   |  |||| : | |  || ||
Db    267  ------ATMVAAGVFLVARMYPVFEPIPEAMNVIAWTGATTAFLGATIALTQNDIKKG  318

Qy    285  VDFTAVRDL  293
                :|::|| |
Db    319  LAYSTMSQL  327
```

Search completed: March 18, 2003, 15:41:03
Job time : 19 secs

```
OM protein - protein search, using sw model

                 GenCore version 5.1.4_p5_4578
                 Copyright (c) 1993 - 2003 Compugen Ltd.

Run on:          March 18, 2003, 15:41:34 ; Search time 14 Seconds
                                             (without alignments)
                                             1224.732 Million cell updates/sec

Title:           1923
Perfect score:   US-09-691-219-2
Sequence:        1 MEAKEKQHLLDARPAIRSYT.........QIEDYSEDMKDCQACQGPSD 372

Scoring table:   BLOSUM62
                 Gapop 10.0 , Gapext 0.5

Searched:        199416 seqs, 46092074 residues

Total number of hits satisfying chosen parameters:    199416

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
                 Maximum Match 100%
                 Listing first 120 summaries

Database :       1: /cgn2_6/ptodata/2/pubpaa/US08_NEW_PUB.pep:*
                 2: /cgn2_6/ptodata/2/pubpaa/PCT_NEW_PUB.pep:*
                 3: /cgn2_6/ptodata/2/pubpaa/US06_NEW_PUB.pep:*
                 4: /cgn2_6/ptodata/2/pubpaa/US06_PUBCOMB.pep:*
                 5: /cgn2_6/ptodata/2/pubpaa/US07_NEW_PUB.pep:*
                 6: /cgn2_6/ptodata/2/pubpaa/US07_PUBCOMB.pep:*
                 7: /cgn2_6/ptodata/2/pubpaa/PCTUS_PUBCOMB.pep:*
                 8: /cgn2_6/ptodata/2/pubpaa/US08_PUBCOMB.pep:*
                 9: /cgn2_6/ptodata/2/pubpaa/US09_NEW_PUB.pep:*
                 10: /cgn2_6/ptodata/2/pubpaa/US09_PUBCOMB.pep:*
                 11: /cgn2_6/ptodata/2/pubpaa/US10_NEW_PUB.pep:*
                 12: /cgn2_6/ptodata/2/pubpaa/US10_PUBCOMB.pep:*
                 13: /cgn2_6/ptodata/2/pubpaa/US60_NEW_PUB.pep:*
                 14: /cgn2_6/ptodata/2/pubpaa/US60_PUBCOMB.pep:*

                 Published Applications AA:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.
```

SUMMARIES

Result No.	Score	Query Match Length	DB	ID	Description
1	902.5	46.9	320	10	US-09-872-153-22	Sequence 22, Appl
2	722.5	37.6	438	10	US-09-957-708-27	Sequence 27, Appl
3	350	18.2	318	9	US-09-738-626-4910	Sequence 4910, Ap
4	347.5	18.1	507	10	US-09-795-927-10	Sequence 10, Appl
5	347.5	18.1	507	10	US-09-738-897-2	Sequence 2, Appli
6	335	17.4	507	10	US-09-738-897-4	Sequence 4, Appli
7	333.5	17.3	490	9	US-10-012-542-158	Sequence 158, App
8	333	17.3	504	9	US-10-012-542-332	Sequence 332, Appl
9	333	17.3	503	10	US-09-738-897-5	Sequence 5, Appli
10	285.5	14.8	564	12	US-10-139-262-28	Sequence 28, Appli
11	278.5	14.5	765	9	US-10-028-072-28	Sequence 28, Appl
12	278.5	14.5	765	9	US-10-121-049-28	Sequence 28, Appl
13	278.5	14.5	765	9	US-10-123-904-28	Sequence 28, Appl
14	278.5	14.5	765	9	US-10-140-470-28	Sequence 28, Appl
15	278.5	14.5	765	9	US-10-175-746-28	Sequence 28, Appl
16	278.5	14.5	765	9	US-10-176-918-28	Sequence 28, Appl
17	278.5	14.5	765	9	US-10-176-921-28	Sequence 28, Appl
18	278.5	14.5	765	9	US-10-137-865-28	Sequence 28, Appl
19	278.5	14.5	765	9	US-10-140-474-28	Sequence 28, Appl
20	278.5	14.5	765	9	US-10-142-431-28	Sequence 28, Appl
21	278.5	14.5	765	9	US-10-143-114-28	Sequence 28, Appl
22	278.5	14.5	765	9	US-10-140-002-28	Sequence 28, Appl
23	278.5	14.5	765	9	US-09-815-242-13500	Sequence 13500, A
24	278.5	14.5	765	9	US-09-738-626-6557	Sequence 6557, Ap
25	272.5	14.2	594	12	US-10-139-262-8	Sequence 8, Appl
26	272.5	14.2	293	9	US-10-139-262-8	Sequence 8, Appli
27	213.5	11.1	293	9	US-10-012-542-233	Sequence 233, App
28	213.5	11.1	307	9	US-10-012-542-331	Sequence 331, App
29	132	6.9	394	10	US-09-815-242-13500	Sequence 13500, A
30	120	6.2	286	9	US-09-738-626-6557	Sequence 6557, Ap
31	100.5	5.2	370	10	US-09-823-114-21	Sequence 21, Appl
32	99.5	5.2	458	9	US-09-738-626-5835	Sequence 5835, Ap
33	94.5	4.9	306	9	US-09-738-626-3686	Sequence 3686, Ap
34	93.5	4.9	372	9	US-09-738-626-5934	Sequence 5934, Ap
35	93	4.8	441	10	US-09-510-332-77	Sequence 77, Appl
36	91	4.7	335	10	US-09-143-634-1	Sequence 1, Appli
37	91	4.7	335	10	US-09-966-871-80	Sequence 80, Appl
38	89.5	4.7	372	10	US-10-035-645-80	Sequence 80, Appl
39	89	4.6	412	10	US-09-815-242-10236	Sequence 10236, A
40	89	4.6	513	10	US-09-738-626-4772	Sequence 4772, Ap
41	88.5	4.6	488	10	US-09-996-634-140	Sequence 140, App
42	85.5	4.4	319	10	US-09-886-055-95	Sequence 95, Appl
43	85.5	4.4	7257	9	US-10-014-717-5	Sequence 5, Appli
44	85	4.4	618	10	US-10-216-355-2	Sequence 2, Appli
45	84	4.4	366	9	US-09-738-626-4530	Sequence 4530, Ap
46	84	4.4	387	9	US-09-738-626-5868	Sequence 5868, Ap
47	84	4.4	387	9	US-09-966-782A-17	Sequence 17, Appl
48	84	4.4	387	10	US-10-212-980-10	Sequence 10, Appl
49	83.5	4.3	518	10	US-09-918-171A-22	Sequence 22, Appl
50	83	4.3	439	10	US-09-815-242-5660	Sequence 5660, Ap
51	83	4.3	484	10	US-09-815-242-12272	Sequence 12272, A
52	82.5	4.3	942	10	US-09-969-528-6	Sequence 6, Appl
53	82	4.3	591	10	US-09-962-290-4	Sequence 4, Appli
54	82	4.3	694	9	US-09-944-413-55	Sequence 55, Appl
55	82	4.3	694	9	US-09-944-403-55	Sequence 55, Appl
56	82	4.3	694	9	US-09-944-896-55	Sequence 55, Appl
57	82	4.3	694	9	US-09-738-626-5934	Sequence 4772, Ap
58	82	4.3	694	9	US-09-944-907-55	Sequence 55, Appl
59	82	4.3	694	9	US-09-944-929-55	Sequence 55, Appl
60	82	4.3	694	9	US-09-944-590-60	Sequence 60, Appl
61	82	4.3	694	9	US-10-176-758-60	Sequence 60, Appl
62	82	4.3	694	9	US-10-175-737-60	Sequence 60, Appl
63	82	4.3	694	9	US-10-173-706-60	Sequence 60, Appl
64	82	4.3	694	9	US-10-175-738-60	Sequence 60, Appl
65	82	4.3	694	9	US-10-175-752-60	Sequence 60, Appl
66	82	4.3	694	9	US-10-176-482-60	Sequence 60, Appl
67	82	4.3	694	9	US-10-176-757-60	Sequence 60, Appl
68	82	4.3	694	9	US-10-176-913-60	Sequence 60, Appl
69	82	4.3	694	9	US-10-180-552-60	Sequence 60, Appl
70	82	4.3	694	9	US-10-180-557-60	Sequence 60, Appl
71	82	4.3	694	9	US-10-173-700-60	Sequence 60, Appl
72	82	4.3	694	9	US-10-174-572-60	Sequence 60, Appl
73	82	4.3	694	9	US-10-174-579-60	Sequence 60, Appl
74	82	4.3	694	9	US-10-174-582-60	Sequence 60, Appl
75	82	4.3	694	9	US-10-174-588-60	Sequence 60, Appl
76	82	4.3	694	9	US-10-175-739-60	Sequence 60, Appl
77	82	4.3	694	9	US-10-175-740-60	Sequence 60, Appl
78	82	4.3	694	9	US-10-175-743-60	Sequence 60, Appl
79	82	4.3	694	9	US-10-176-488-60	Sequence 60, Appl
80	82	4.3	694	9	US-10-176-492-60	Sequence 60, Appl
81	82	4.3	694	9	US-10-176-747-60	Sequence 60, Appl
82	82	4.3	694	9	US-10-176-750-60	Sequence 60, Appl
83	82	4.3	694	9	US-10-176-985-60	Sequence 60, Appl
84	82	4.3	694	9	US-10-176-987-60	Sequence 60, Appl
85	82	4.3	694	9	US-10-176-991-60	Sequence 60, Appl
86	82	4.3	694	9	US-10-176-992-60	Sequence 60, Appl
87	82	4.3	694	9	US-10-176-993-60	Sequence 60, Appl
88	82	4.3	694	9	US-10-184-658-60	Sequence 60, Appl
89	82	4.3	694	9	US-10-173-695-60	Sequence 60, Appl
90	82	4.3	694	9	US-10-173-697-60	Sequence 60, Appl
91	82	4.3	694	9	US-10-173-705-60	Sequence 60, Appl
92	82	4.3	694	9	US-10-174-576-60	Sequence 60, Appl

```
 93  82  4.3  694  9  US-10-174-585-60  Sequence 60, Appl
 94  82  4.3  694  9  US-10-174-586-60  Sequence 60, Appl
 95  82  4.3  694  9  US-10-175-747-60  Sequence 60, Appl
 96  82  4.3  694  9  US-10-184-481-60  Sequence 60, Appl
 97  82  4.3  694  9  US-10-184-485-60  Sequence 60, Appl
 98  82  4.3  694  9  US-10-176-487-60  Sequence 60, Appl
 99  82  4.3  694  9  US-10-176-493-60  Sequence 60, Appl
100  82  4.3  694  9  US-10-176-756-60  Sequence 60, Appl
101  82  4.3  694  9  US-10-176-911-60  Sequence 60, Appl
102  82  4.3  694  9  US-10-176-919-60  Sequence 60, Appl
103  82  4.3  694  9  US-10-176-925-60  Sequence 60, Appl
104  82  4.3  694  9  US-10-176-978-60  Sequence 60, Appl
105  82  4.3  694  9  US-10-179-510-60  Sequence 60, Appl
106  82  4.3  694  9  US-10-180-543-60  Sequence 60, Appl
107  82  4.3  694  9  US-10-180-544-60  Sequence 60, Appl
108  82  4.3  694  9  US-10-180-546-60  Sequence 60, Appl
109  82  4.3  694  9  US-10-180-547-60  Sequence 60, Appl
110  82  4.3  694  9  US-10-180-549-60  Sequence 60, Appl
111  82  4.3  694  9  US-10-180-555-60  Sequence 60, Appl
112  82  4.3  694  9  US-10-180-559-60  Sequence 60, Appl
113  82  4.3  694  9  US-10-181-000-60  Sequence 60, Appl
114  82  4.3  694  9  US-10-183-010-60  Sequence 60, Appl
115  82  4.3  694  9  US-10-183-012-60  Sequence 60, Appl
116  82  4.3  694  9  US-10-184-614-60  Sequence 60, Appl
117  82  4.3  694  9  US-10-184-623-60  Sequence 60, Appl
118  82  4.3  694  9  US-10-184-635-60  Sequence 60, Appl
119  82  4.3  694  9  US-10-184-637-60  Sequence 60, Appl
120  82  4.3  694  9  US-10-184-646-60  Sequence 60, Appl
```

ALIGNMENTS

RESULT 1
US-09-872-153-22
Sequence 22, Application US/09872153
Patent No. US20020082207A1
GENERAL INFORMATION:
APPLICANT: Hirst, Shannon K.
APPLICANT: Harlocker, Susan L.
APPLICANT: Dillon, Davin C.
APPLICANT: Kalos, Michael D.
TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY
TITLE OF INVENTION: AND DIAGNOSIS OF PANCREATIC CANCER
FILE REFERENCE: 210121.531
CURRENT APPLICATION NUMBER: US/09/872,153
CURRENT FILING DATE: 2001-05-31
NUMBER OF SEQ ID NOS: 28
SOFTWARE: FastSEQ for Windows Version 4.0
SEQ ID NO 22
LENGTH: 320
TYPE: PRT
ORGANISM: Homo sapiens
US-09-872-153-22

```
Query Match        46.9%;  Score 902.5;  DB 10;  Length 320;
Best Local Similarity  53.1%;  Pred. No. 1.4e-83;
Matches  171;  Conservative  61;  Mismatches  87;  Indels  3;  Gaps  1;

Qy   51  HHCHAQKGPDSHCDRKKGKAQRQLYVASAICLLFMIGEVVGGYLAHSLAVMTDAAHLLTD  110
          :!!:!                     !!! |   !!! . | !!!|!|:|!|!!!!!!!
Db    2  YHCHSGGSKPTEKGANEYAYAKWKLCSASAICPIFMIAEVVGGHIAGSLAVVTDAAHLLID   61

Qy  111  FASMLISLFSLWMSSRPATKTMNFGWQRAEILGALVLVSLSIWVTGVLVLYLAVERLISGD  170
          :!||||||||!|:| ||||||!:|! :!!||||||||||||||!||| !:!!||||!|
Db   62  LTSFLLSLFSLWLSKRPPSKRLTFGWHRAEILGALLSILCIWVVTGVLVLVYLACERLLYPD  121

Qy  171  YEIDGGTMLITSGCAVAVNIIMGLTLHQSGHGHSHGTTNQQEENPSVRAAFIHVIGDEMQ  230
          |:!                |:| |||! !||:.. !!|!! |  | ||||||||||:||
Db  122  YQIQATVMIIVSSCAVAANIVLTVVLHQRCLGHNH---KEVQANASVRAAFVHALGDLFQ  178

Qy  231  SMGVLIVAAYILYFKPEYKYVDPICTFVFSILVLVGTTLTILRDVILVLMEGTPKGVDFTAV  290
          |:|!||!!:|!||!!!! !!|!!!||!|!!||!|||!!!:!!!!!!|||!|||
Db  122  YQIQATVMIIVSSCAVAANIVLTVVLHQRCLGHNH---KEVQANASVRAAFVHALGDLFQ  178
```

RESULT 2
US-09-957-708-27
Sequence 27, Application US/09957708
Publication No. US20030031678A1
GENERAL INFORMATION:
APPLICANT: Sun, Yongming
APPLICANT: Recipon, Herve
APPLICANT: Cafferkey, Robert
APPLICANT: Ali, Shujath
TITLE OF INVENTION: Compositions and Methods Relating to Prostate Specific
TITLE OF INVENTION: Genes
FILE REFERENCE: DEX-0239
CURRENT APPLICATION NUMBER: US/09/957,708
CURRENT FILING DATE: 2001-09-19
PRIOR APPLICATION NUMBER: 60/233,746
PRIOR FILING DATE: 2000-09-19
NUMBER OF SEQ ID NOS: 40
SOFTWARE: Patentin Ver. 2.1
SEQ ID NO 27
LENGTH: 438
TYPE: PRT
ORGANISM: Homo sapiens
US-09-957-708-27

```
Query Match        37.6%;  Score 722.5;  DB 9;  Length 438;
Best Local Similarity  39.0%;  Pred. No. 3.9e-65;
Matches  151;  Conservative  88;  Mismatches  117;  Indels  31;  Gaps  7;

Qy    2  EAKEKQHLLDARPAIRSYTGSLWQEGAGWIPLPRPGLDLQAIELAAQSNHHCHAQKGPDS   61
          !! :|!     !| ||:||: ||!  |!  .  .|!:|| |! ! |  : :| ||||||
Db   56  EAPERP-VNGAHPTLQADDDSLLDQD--LPLINSQLSLK-----VDSCDNCSKQRE---  103

Qy   62  HCDPKKGKAQRQLYVASAICLLFMIGEVVGGYLAHSLAVMTDAAHLLTDFASMLISLFSL  121
           |  | !  !!|||:|||||||!| :  !!|||||||||!|||!||||!|!|:|!:|||
Db  104  --ILKQRVKARLTIAAVLYLFMIGELVGGYIANSLAIMTDALHMLTDLSALILTLLAL  161

Qy  122  WMSSRPATKTMNFGWQRAEILGALVSVLSIWVVTGVLVLYLAVERLISGDYEIDGGTMLIT  181
          |!!|:   | |!|||:||!|||| !!|| |!  |||!|!  :!||!|!|| | :|| !!!
Db  162  WLSSKSPTKRFTFGFHRLEVLSAMISVLLVYILMGFLLYEAVQRTIHMNYEINGDIMLIT  221

Qy  182  SGCAVAVNIIMGLTLHQSGHGHSHGTT----------------NQQEENPSVRAAFIHVIGD  227
          ||||:|!|!!|||!||| |||!! | !           :: :|!|||!|! !!
Db  222  AAVGVAVNVIMGFLLNQSGHRHSHSHSLPSNSPTRGSGCERNHGQDSLAVRAAFVHALGD  281

Qy  228  FMQSMGVLVAAYILYFKPEYKYVDPICTFVFSILVLVGTTLTILRDVILVLMEGTPKGVDF  287
          | ! !!| |||!||||!|||:|!| :||!!|!!|| |! :!!|!|:!||||||!|  |
Db  282  LVQSVGVLIAAYIIRFKPEYKIADPICTYVFSLLVAFTFRIIWDTVVIILEGVPSHLNV  341

Qy  288  TAVRDLLLSVEGVEALHSLHIWALTVAQPVLSVHIAIAQNTDA--QAVLKTASSRLQGKF  345
          ||:||!|!|||!||||||!||!|||||!|||||||||!||||   !!|||||||||||!
Db  342  DYIKEALMKIEDVYSVEDLNIWSLTSGKSTAIVHIQLIPGSSSKWEEVQSKANHLLLNTF  401

Qy  346  HFHIVTIQIEDYSEDM-KDCQACQGPS  371
           |!|||!!| :|!|! !||||||||||
Db  402  GMYRCTIQLQSYRQEVDRTCANCQSSS  428
```

RESULT 3
US-09-738-626-4910
Sequence 4910, Application US/09738626
Publication No. US20020197605A1
GENERAL INFORMATION:

APPLICANT: NAKAGAWA, SATOSHI
APPLICANT: MIZOGUCHI, HIROSHI
APPLICANT: ANDO, SEIKO
APPLICANT: HAYASHI, MIKIRO
APPLICANT: OCHIAI, KEIKO
APPLICANT: YOKOI, HARUHIKO
APPLICANT: TATEISHI, NAOKO
APPLICANT: SENOH, AKIHIRO
APPLICANT: IKEDA, MASATO
APPLICANT: OZAKI, AKIO
TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
FILE REFERENCE: 249-125
CURRENT APPLICATION NUMBER: US/09/738,626
CURRENT FILING DATE: 2000-12-18
PRIOR APPLICATION NUMBER: JP 99/377484
PRIOR FILING DATE: 1999-12-16
PRIOR APPLICATION NUMBER: JP 00/159162
PRIOR FILING DATE: 2000-04-07
PRIOR APPLICATION NUMBER: JP 00/280988
PRIOR FILING DATE: 2000-08-03
NUMBER OF SEQ ID NOS: 7059
SOFTWARE: Patentin ver. 3.0
SEQ ID NO 4910
LENGTH: 318
TYPE: PRT
ORGANISM: Corynebacterium glutamicum
US-09-738-626-4910

Query Match   18.2%;   Score 350;   DB 9;   Length 318;
Best Local Similarity   27.5%;   Pred. No. 1.5e-27;
Matches 92;   Conservative 72;   Mismatches 141;   Indels 30;   Gaps 9;

```
Qy    39   DLQAIELAAQSNHHCHAQKGPDSHCDPKKGKAQORQLYVASAICLLFMIGEVVGGYLAHSL    98
                 |||  |:|  |:||          |    | |||:||||      |:|
Db     5   DLQQRSYAHNPDGHDHSHDGLGHSHAPSSLKA---LFAVIIFTSIIFLAELIAGLISGSL    61

Qy    99   AVMTDAAHLLTDFASMLISLFSLWMSSRPATKTMNFGWQRAEILGALVSVLSIWVVTGVL   158
                 || |        ||  |    ||||         |     |||      ||||
Db    62   ALLADAMHMLSDSTGLIIAAVAMLIGRRARTSRATGYKRAEVLAAMVNAT---VVTALS   118

Qy   159   VYLAVERL--ISGDYEIDGGTMLITSGCAVAVNIIMGLTL--HQSGHGHSHGTTNQQEEN   214
                 |:||  |    |  || |              ||||   |   |
Db   119   VWIVVEAIMRLGKDLEIQTNLMLIVAVIGFVTNGISALVLMRHQDG----------N   165

Qy   215   PSVRAAFIHVIGDFMQSMGVLVAAYILLYFKPEYKVDPICTEVFSILVLGTTLTILRDVI   274
                 | |:|| | ::| |  |  |  |:|| |:  |:|||:  ||:  ||| |  |::|
Db   166   INMRGAFLHVLSDMLGSVAVIIAGLVIRY-TGWMPADTIASIAIAAIIIPRAFSLKEAL   224

Qy   275   LVIMEGTPKGVDFTAVRDLLLSVEGVEALHSLHIWALTVAQPVLSVHIAIAQNTD---AQ   331
                 | || | ||| :||| ||| |:|||| | || ||||:|  | |  |       |
Db   225   NILLERVPTGAEPAEVDAALRKVPGVSDVHDLHIWSIDGKEILATVHLVVDSSTNQLHSC   284

Qy   332   AVLKTASSRLQGKFHFHTVTIQIE--DYSEDMKDC   364
                 | |:|||  | |  |:|: ||
Db   285   GVLDRAEAELSKLGILHS-TIQLESADHSDHESVC   318
```

RESULT 4
US-09-795-927-10
Sequence 10, Application US/09795927
Publication No. US20030039962A1
GENERAL INFORMATION:
APPLICANT: Walke, D. Wade
APPLICANT: Wilganowski, Nathaniel L.
APPLICANT: Hu, Yi
APPLICANT: Kieke, James A.
APPLICANT: Zambrowicz, Brian
APPLICANT: Revelli, Jean-Pierre
APPLICANT: Turner, C. Alexander Jr.
TITLE OF INVENTION: NO. US20030039962A1le1 Human Transporter Proteins and Polynucleot:
TITLE OF INVENTION: Same
FILE REFERENCE: LEX-0141-USA
CURRENT APPLICATION NUMBER: US/09/795,927

CURRENT FILING DATE: 2001-02-28
PRIOR APPLICATION NUMBER: US 60/185,956
PRIOR FILING DATE: 2000-02-29
NUMBER OF SEQ ID NOS: 11
SOFTWARE: FastSEQ for Windows Version 4.0
SEQ ID NO 10
LENGTH: 507
TYPE: PRT
ORGANISM: homo sapiens
US-09-795-927-10

Query Match   18.1%;   Score 347.5;   DB 9;   Length 507;
Best Local Similarity   24.6%;   Pred. No. 5.1e-27;
Matches 108;   Conservative 71;   Mismatches 117;   Indels 143;   Gaps 13;

```
Qy    68   GKAQORQLYVASAICLLFMIGEVVGGYLAHSLAVMTDAAHLLTDFASMLISLFSLWMSSRP   127
                 |:|: | | :||          |    | |||:||||      |:|
Db     5   GRNRGRLLCMLALFTFMFMVLEVVVSRVTSSLAMLSDSFHMLSDVLALVVALVAERFARRT    64

Qy   128   -ATKTMNFGWQRAEILGALVSVLSIWVVTGVLVLVLAVERLISGDYEIDGGTMLITSGCAV   186
                 |:    | |||:||| | :|||:|:: | |||  |||::| | |||
Db    65   HATQKNTFGWIRAEVMGALVNAIFLTGLCFAILLEAIERFIE-PHEMQQPLVVLGVGVAG   123

Qy   187   -------HQSGHGHSHG--------   206
                 |||||||
Db   124   LLVNVLGLCLFHHHSGFSQDSGHGHSHGHGHGHGLPKGPRVKSTRPGSSDINVAPGEQG   183

Qy   207   -----------TTNQ------QEENP---------   215
                 :|
Db   184   PDQEETNTLVANTSNSGNLKLDPADPENPRSGDTVEVQVNGNLVREPDHMELEDRAGQL   243

Qy   216   SVRAAFIHVIGDFMQSMGVLVAAYILYFK------PE------   246
                 ::|     || ||         |  |           |::
Db   244   NMRGVFLHVLGDALGSVIVVVNALVFYFSWKGCSEGDFCVNPCFPDPCKAFVEIINSTHA   303

Qy   247   --YK-------YVDPICTEVFSILVLGTTLTILRDVILVLMEGTPKGVDFTAVRDLLLSV   297
                 |:     |||   ||: ||:  ||||  | : ||
Db   304   SLYEAGPCWVLYLDPTLCVVMVCILLLYTTYPLLKESALILLQTVPKQIDIRNLIKELRNV   363

Qy   298   EGVEALHSLHIWALTVAQPVLSVHIAIAQNTDAQAVLKTASSRLQGKFH---FHTVTIQI   354
                 |:|||| | || ||||:| | |  |         ||  ||||  | |
Db   364   EGVEEVHELHVWQLAGSRIIATAHIKCEDPTSYMEVAKT----IKDVFHHNGIHATTIQP   419

Qy   355   EDYSEDMKD-----CQ-ACQ  368
                 :| | ||:      |: ||:
Db   420   EFASVGSKSSVVPCELACR  438
```

RESULT 5
US-09-738-897-2
Sequence 2, Application US/09738897
Patent No. US20020106772lA1
GENERAL INFORMATION:
APPLICANT: GUEGLER, Karl et al
TITLE OF INVENTION: ISOLATED HUMAN TRANSPORTER PROTEINS,
TITLE OF INVENTION: NUCLEIC ACID MOLECULES ENCODING HUMAN TRANSPORTER PROTEINS,
TITLE OF INVENTION: AND USES THEREOF
FILE REFERENCE: CL001053
CURRENT APPLICATION NUMBER: US/09/738,897
CURRENT FILING DATE: 2000-12-18
NUMBER OF SEQ ID NOS: 5
SOFTWARE: FastSEQ for Windows Version 4.0
SEQ ID NO 2
LENGTH: 507
TYPE: PRT
ORGANISM: Human
US-09-738-897-2

Query Match   18.1%;   Score 347.5;   DB 10;   Length 507;
Best Local Similarity   24.6%;   Pred. No. 5.1e-27;
Matches 108;   Conservative 71;   Mismatches 117;   Indels 143;   Gaps 13;

```
Qy    68   GKAQORQLYVASAICLLFMIGEVVGGYLAHSLAVMTDAAHLLTDFASMLISLFSLWMSSRP   127
```

RESULT 6
US-09-738-897-4

Sequence 4, Application US/09738897
; Patent No. US20020106721A1
GENERAL INFORMATION:
APPLICANT: GUEGLER, Karl et al
TITLE OF INVENTION: ISOLATED HUMAN TRANSPORTER PROTEINS,
TITLE OF INVENTION: NUCLEIC ACID MOLECULES ENCODING HUMAN TRANSPORTER PROTEINS,
TITLE OF INVENTION: AND USES THEREOF
FILE REFERENCE: CL001053
CURRENT APPLICATION NUMBER: US/09/738,897
CURRENT FILING DATE: 2000-12-18
NUMBER OF SEQ ID NOS: 5
SOFTWARE: FastSEQ for Windows Version 4.0
SEQ ID NO 4
LENGTH: 507
TYPE: PRT
ORGANISM: Rattus norvegicus
US-09-738-897-4

```
Query Match            17.4%;  Score 335;  DB 10;  Length 507;
Best Local Similarity  24.0%;  Pred. No. 9.5e-26;
Matches   105;  Conservative 73;  Mismatches 118;  Indels 142;  Gaps 13;

Db    5 GRNRGRLLCMLLTFMFMVLEVVSRVTSSLAMLSDSFHMLSDVLALVVALVAERFARRT 64
          :|| |||  :|:   :|  |||:|| :|    ::|:|
Qy  128 -ATKTMNFGWQRAEILGALVSVLSIWVVTGVLVLVYLAVERLISGDYEIDGGTMLIITSGCAV 186

Db   65 HATQKNTFGWIRAEVMGALVNAIFLTGLCFAILLEAIERFIE-PHEMQQPLVVLGVGVAG 123
           ||||||||:|||:| |:| :|
Qy  187 ------HQSGHGHSHG------ 206

Db  124 LLIVNVLGLCLFHHHSGFSQDSGHGHSGHGHGHGHGLPKGPRVKSTRPGSSSDINVAPGEQG 183
            |||:| :::|
Qy  207 -------QEENP------ 215

Db  184 PDQQEETNTLVANTSNSNGLKLDPADPENRSGDTVEVQVQVNGNLVREPDHMELEEDRAGQL 243
            :|   |:
Qy  216 SVRAAFIHVIGDFMQSMGVLVAAYILYFK 246

Db  244 NMRGVFLHVLGDALGSVIVVVNALVFYFSWKGCSEGFDCVNPCFPDPCKRAFVEIINSTHA 303
            ||||          -PE-
Qy  247 --YK-------YVDPICTFVFSILVLGTTLILTILRDVILVLMEGTPKGVDFTAVRDLLSV 297

Db  304 SVFEAGPCWVLYLDPTLCVVMVCILLYTTYPLLKESALILLQTVPKQIDIRNLIKELRNV 363
             |:
Qy  298 EGVEALHSLHIWALTVAQPVLSVHIAIAQNTDAQAVLKTASSRLQGKFH---FHTVTIQI 354

Db  364 EGVEEVHELHVWQLAGSRIIATAHKCEDPTSYMEVAKT-----IKDVFHNHGIHATTIQP 419
            :|:
Qy  355 EDYSEDMKD----CQ-ACQ 368

Db  420 EFASVGSKSSVVPCELACR 438
```

RESULT 7
US-10-012-542-158

Sequence 158, Application US/10012542
; Publication No. US20030044851A1
GENERAL INFORMATION:
APPLICANT: Ruben et al.
TITLE OF INVENTION: 94 Human Secreted Proteins
FILE REFERENCE: PF029P1
CURRENT APPLICATION NUMBER: US/10/012,542
CURRENT FILING DATE: 2001-12-12
PRIOR APPLICATION NUMBER: 09/461,325
PRIOR FILING DATE: 1999-12-14
PRIOR APPLICATION NUMBER: 60/089,509
PRIOR FILING DATE: 1998-06-16
PRIOR APPLICATION NUMBER: 60/089,510
PRIOR FILING DATE: 1998-06-16
PRIOR APPLICATION NUMBER: 60/089,507
PRIOR FILING DATE: 1998-06-16
PRIOR APPLICATION NUMBER: 60/089,508
PRIOR FILING DATE: 1998-06-16
PRIOR APPLICATION NUMBER: 60/090,112
PRIOR FILING DATE: 1998-06-22
PRIOR APPLICATION NUMBER: 60/090,113
PRIOR FILING DATE: 1998-06-22
NUMBER OF SEQ ID NOS: 532
SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 158
LENGTH: 490
TYPE: PRT
ORGANISM: Homo sapiens
FEATURE:
NAME/KEY: SITE
LOCATION: (134)
OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
NAME/KEY: SITE
LOCATION: (389)
OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
US-10-012-542-158

```
Query Match            17.3%;  Score 333.5;  DB 9;  Length 490;
Best Local Similarity  24.5%;  Pred. No. 1.3e-25;
Matches   105;  Conservative 67;  Mismatches 113;  Indels 143;  Gaps 13;

Db  244 MRGVFLHVLGDALGSVIVVVNALVFYFSWKGCTEDDCVNPCFPDPCKSSVELMNSTQAP 303
            ||||
Qy  246 EYK-------YVDPICTFVFSILVLGTTLILTILRDVILVLMEGTPKGVDFTAVRDLLSVE 298

Db  304 MHEAGPCWVLYLDPTLCIIMVCILLYTTYPLLKESALILLQTVPKQIDIKHLVKELRDVE 363
Qy  299 GVEALHSLHIWALTVAQPVLSVHIAIAQNTDAQAVLKTASSRLQGKFH---FHTVTIQIE 355

Db  364 GVEEVHELHVWQLAGSRIIATAHKCEDPTASYMQVAKT-----IKDVFHNHGIHATTIQPE 419
Qy  356 DYSEDMKD----CQ-ACQ 368

Db  420 FASVGSKSSVVPCELACR 437
```

```
Db  182 NTSNSNGLKLDPADPENRSGDTVEVQVQNGNLVREPDHMELEDRAGQLMRGVFLHVLG 241

Qy  227 DFMQSMGVLVAAYILYF----------------------KPEYK------------ 248

Db  242 DALGSVIVVVNALVFYFSWKGCSEGDFCVNPCGFPDPCKPFVEIINSTHASVYEAGPCWVL 301

Qy  249 YVDPICTFVFSILVLGTTFLTILRDVILVLMEGTPKGVDFTAVRDLLLSVEGVEALHSLHI 308

Db  302 YLDPTLCVVMVCILLYTTYPLLKESALILLQTVPKQIDIRNLIKELRNVEGVEEVHELHV 361

Qy  309 WALTVAQPVLSVHIAIAQNTDAQAVLKTASSRLQGKFH---FHTVTIQIEDYSEDMKD-- 363

Db  362 WQLAGSRIIATAHIKCEDPTSYMEVAKX----IKDVFHNHGIHATTIQPEFASVGSKSSV 417

Qy  364 --CQ-ACQ 368

Db  418 VPCELACR 425
```

RESULT 8
US-10-012-542-332
; Sequence 332, Application US/10012542
; Publication No. US20030048851A1
; GENERAL INFORMATION:
; APPLICANT: Ruben et al.
; TITLE OF INVENTION: 94 Human Secreted Proteins
; FILE REFERENCE: PZ029P1
; CURRENT APPLICATION NUMBER: US/10/012,542
; CURRENT FILING DATE: 2001-12-12
; PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: 09/461,325
; PRIOR FILING DATE: EARLIER FILING DATE: 1999-12-14
; PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: 60/089,507
; PRIOR FILING DATE: EARLIER FILING DATE: 1998-06-16
; PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: 60/089,508
; PRIOR FILING DATE: EARLIER FILING DATE: 1998-06-16
; PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: 60/089,509
; PRIOR FILING DATE: EARLIER FILING DATE: 1998-06-16
; PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: 60/089,510
; PRIOR FILING DATE: EARLIER FILING DATE: 1998-06-16
; PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: 60/090,112
; PRIOR FILING DATE: EARLIER FILING DATE: 1998-06-22
; PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: 60/090,113
; PRIOR FILING DATE: EARLIER FILING DATE: 1998-06-22
; SOFTWARE: Patentin Ver. 2.0
; NUMBER OF SEQ ID NOS: 532
; SEQ ID NO 332
; LENGTH: 504
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: SITE
; LOCATION: (148)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
; NAME/KEY: SITE
; LOCATION: (403)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
US-10-012-542-332

Query Match 17.3%; Score 333.5; DB 9; Length 504;
Best Local Similarity 24.5%; Pred. No. 1.3e-25;
Matches 105; Conservative 67; Mismatches 113; Indels 143; Gaps 13;

```
Qy   79 AICLLFMIGEVVGGYLAHSLAVNMTDAAHLLTDFASMLISLFSLWMSSRP-ATKTMNFGWQ 137

Db   17 ALTFMFWVLEVVVSRVTSSLAMLSDSFHMLSDVLALVVALVAERFARTHATQKNTFGWI 76

Qy  138 RAEILGALVSVLSIWVVTGVLVYLAVERLISGDYEIDGGTMLITSGCAVAVNIIMGLTL- 196

Db   77 RAEVMGALVNAIFLTGLCFAILLEAIERFIE-PHEMQQPLVVLGVGVAGLLVNVLGLCLF 135

Qy  197 -HQS--------------GHGHSHG------------- 206

Db  136 HHHSGFSQDSGHXHSHGHGHGHLPKGPRVKSTRPGSSDINVAPGEQGPDQEETNTLVA 195
```

RESULT 9
US-09-738-897-5
; Sequence 5, Application US/09738897
; Patent No. US20020106721A1
; GENERAL INFORMATION:
; APPLICANT: GUEGLER, Karl et al
; TITLE OF INVENTION: ISOLATED HUMAN TRANSPORTER PROTEINS,
; TITLE OF INVENTION: NUCLEIC ACID MOLECULES ENCODING HUMAN TRANSPORTER PROTEINS,
; TITLE OF INVENTION: AND USES THEREOF
; FILE REFERENCE: CL001053
; CURRENT APPLICATION NUMBER: US/09/738,897
; CURRENT FILING DATE: 2000-12-18
; SOFTWARE: FastSEQ for Windows Version 4.0
; NUMBER OF SEQ ID NOS: 5
; SEQ ID NO 5
; LENGTH: 503
; TYPE: PRT
; ORGANISM: Mus musculus
US-09-738-897-5

Query Match 17.3%; Score 333; DB 10; Length 503;
Best Local Similarity 23.7%; Pred. No. 1.5e-25;
Matches 103; Conservative 72; Mismatches 121; Indels 138; Gaps 12;

```
Qy   68 GKAQRQLYVASAICLLFMIGEVVGGYLAHSLAVMTDAAHLLTDFASMLISLFSLWMSSRP 127

Db    5 GRNRGRLLCMLLLTFEMFVLEVVVSRVTASIAMLSDSFHMLSDVLALVVALVAERFARRT 64

Qy  128 -ATKTMNFGWQRAELLGALVSVLSIWVVTGVLVYLAVERLISGDYEIDGGTMLITSGGCAV 186

Db   65 HATQKNTFGWIRAEVMGALVNAIFLTGLCFAILLEAVERFIE-PHEMQQPLVVLSVGVAG 123

Qy  187 AVNIIMGLTL--HQS------GHGHSHG------------ 212

Db  124 LLVNVLGICLFHHHSGEGQGAGHGHSHGHGHGHLAKGARKAGRACGVEAGAPPGRAPDQEE 183

Qy  213 ENP-----------------------------------SVRAA 220

Db  184 TNTLVANTSNSNGLKADQAEPEKLRSDPPVDVQVNGNLIQESDNLEAEDNRAGQLNMRGV 243

Qy  221 FIHVIGDFMQSMGVLVAAYILYFK--------------PE------------ 246

Db  244 FLHVLGDALGSVIVVVNALVFYFNWKGCTEDDFCTNPCFPDPCKSSVEIINSTQAPMRDA 303

Qy  247 ----YKYVDPICTFVFSILVLGTTFLTILRDVILVLMEGTPKGVDFTAVRDLLLSVEGVEA 302

Db  304 GPCWLYLDPTLCIIMVCILLYTTYPLLKESALILLQTVPKQIDIKHLVKELRDVDGVEE 363

Qy  303 LHSLHIWALTVAQPVLSVHIAIAQNTDAQAVLKTASSRLQGKFH---FHTVTIQIEDYSE 359

Db  364 VHELHVWQLAGSRIIATAHIKCEDPASYMQVAKT----IKDVFHNHGIHATTIQPEFASV 419
```

```
Qy  207 -TTNQ--------QEENP-----------------------SVRAAFIHVIG 226

Db  196 NTSNSNGLKLDPADPENRSGDTVEVQVQNGNLVREPDHMELEDRAGQLNMRGVFLHVLG 255

Qy  227 DFMQSMGVLVAAYILYF----------------------KPEYK------------ 248

Db  256 DALGSVIVVVNALVFYFSWKGCSEGDFCVNPCFPDPCKPFVEIINSTHASVYEAGPCWVL 315

Qy  249 YVDPICTFVFSILVLGTTFLTILRDVILVLMEGTPKGVDFTAVRDLLLSVEGVEALHSLHI 308

Db  316 YLDPTLCVVMVCILLYTTYPLLKESALILLQTVPKQIDIRNLIKELRNVEGVEEVHELHV 375

Qy  309 WALTVAQPVLSVHIAIAQNTDAQAVLKTASSRLQGKFH---FHTVTIQIEDYSEDMKD-- 363

Db  376 WQLAGSRIIATAHIKCEDPTSYMEVAKX----IKDVFHNHGIHATTIQPEFASVGSKSSV 431

Qy  364 --CQ-ACQ 368

Db  432 VPCELACR 439
```

```
Qy    360  DMKD-----CQ-ACQ 368
                |:|| |:  |:
Db    420  GSKSSVLPCELACR 433
```

RESULT 10
US-10-139-262-28
; Sequence 28, Application US/10139262
Db Patent No. US2002012845 9A1
GENERAL INFORMATION:
;   APPLICANT: NAKAMURA, YUSUKE
;   APPLICANT: TANAKA, TOSHIHIRO
;   APPLICANT: TSUKADA, SHUICHI
;   TITLE OF INVENTION: PHYSIOLOGICALLY ACTIVE PROTEIN ORIGINATING IN MAMMALS
;   FILE REFERENCE:
;   CURRENT APPLICATION NUMBER: US/10/139,262
;   CURRENT FILING DATE: 2002-05-07
;   PRIOR APPLICATION NUMBER: US/09/380,287A
;   PRIOR FILING DATE: 1999-11-16
;   PRIOR APPLICATION NUMBER: JP   1997-62259
;   PRIOR FILING DATE: 1997-02-28
;   PRIOR APPLICATION NUMBER: JP   1998-62263
;   PRIOR FILING DATE: 1998-02-25
;   NUMBER OF SEQ ID NOS: 42
;   SEQ ID NO 28
;   SOFTWARE: Patentin Ver. 2.1
US-10-139-262-28
;   LENGTH: 564
;   TYPE: PRT
;   ORGANISM: Mus musculus

Query Match   14.8%;   Score 285.5; DB 12;   Length 564;
Best Local Similarity   24.4%;   Pred. No. 1.2e-20;
Matches   77; Conservative 76; Mismatches 124; Indels   39; Gaps   7;

```
Qy     72  RQLYVASAICLLFMIGEVVGGYLAHSLAVMTDAAHLLTDFASMLISLFSLMMSSRPATKT 131
                :  |||    | || :   ||||  : ||||| :   |||    |  |:| :  | : ||
Db    246  RQIYFLCLNLLFTEVELFYGVLTNSLGLISDGFHMLFDCSALVMGLFAALMSRWKATRI 305

Qy    132  MNFGWQRAEILGALVSVLSIWVVTGVVLVAVERLISGDYEIDGGTMLITSGCAVAVNII 191
                ::| | |||     |:|  |:|:|  |   :  |||     |||:| |     |:|:|
Db    306  FSYGYGRIEILSGFINGLFLIVIAFFVMESVARLIDPP-ELDTNMLTPVSVAGLIVNLI 364

Qy    192  ------------------MGLTLHQSGHGHSHGTIN---QQEENPSVRAAFIHVI 225
                                  :||  :: : :::|     : || | :::::::| |
Db    365  GICAFSHAHSHGHGASOGNCHSDHGHSHHAHGHGHDHGHSHGFTGGGMNANMRGVFLHVL 424

Qy    226  GDFMQSMGVLVAAYLIYFKPEYKYVDPICTFVFSIILVLGTTILTILRDVIILVLMEGTPKGV 285
                |:|:|||:|   :  ||  :   ||||| |:|::| |   |  ||:|  ||:||| |:
Db    425  ADTLGSIGVIVST-VLIEQFGWFIADPLCSLFIAVLIFLSVIPLIKDACQVLLRLPPDH 483

Qy    286  DFTAVRDLLLSVEGVEALHSL------HIWAITVAQPVLSVHIAIRQNTDAQAVLKTASS 339
                 ::| |||||    |:  |             : ||  |    |  ||||    | :
Db    484  E---KELHIALEKIQKIEGLISYRDPHFWRHSASIVAGTIHIQVTSEVLEQRIVQQVTG 539

Qy    340  RLQGKFHFHIVTIQIE 355
                 |:     :  |||:||:|
Db    540  ILKDA-GVNNLTIQVE 554
```

RESULT 11
US-10-028-072-28
; Sequence 28, Application US/10028072
Db Publication No. US2003004311A1
GENERAL INFORMATION:
;   APPLICANT: Baker, Kevin P.
;   APPLICANT: Beresini, Maureen
;   APPLICANT: DeForge, Laura
;   APPLICANT: Desnoyers, Luc
;   APPLICANT: Filvaroff, Ellen
;   APPLICANT: Gao, Wei-Qiang
;   APPLICANT: Gerritsen, Mary E.

;   APPLICANT: Goddard, Audrey
;   APPLICANT: Godowski, Paul J.
;   APPLICANT: Gurney, Austin L.
;   APPLICANT: Sherwood, Steven
;   APPLICANT: Smith, Victoria
;   APPLICANT: Stewart, Timothy A.
;   APPLICANT: Tumas, Daniel
;   APPLICANT: Watanabe, Colin K
;   APPLICANT: Wood, William
;   APPLICANT: Zhang, Zemin
;   TITLE OF INVENTION:
;   FILE REFERENCE:
;   CURRENT APPLICATION NUMBER: US/10/028,072
;   CURRENT FILING DATE: 2001-12-19
;   PRIOR APPLICATION NUMBER: 60/049911
;   PRIOR FILING DATE: 1997-06-18
;   PRIOR APPLICATION NUMBER: 60/056974
;   PRIOR FILING DATE: 1997-08-26
;   PRIOR APPLICATION NUMBER: 60/059113
;   PRIOR FILING DATE: 1997-09-17
;   PRIOR APPLICATION NUMBER: 60/059115
;   PRIOR FILING DATE: 1997-09-17
;   PRIOR APPLICATION NUMBER: 60/059117
;   PRIOR FILING DATE: 1997-09-17
;   PRIOR APPLICATION NUMBER: 60/059122
;   PRIOR FILING DATE: 1997-09-17
;   PRIOR APPLICATION NUMBER: 60/059184
;   PRIOR FILING DATE: 1997-09-17
;   PRIOR APPLICATION NUMBER: 60/059263
;   PRIOR FILING DATE: 1997-09-18
;   PRIOR APPLICATION NUMBER: 60/059352
;   PRIOR FILING DATE: 1997-09-19
;   PRIOR APPLICATION NUMBER: 60/059588
;   PRIOR FILING DATE: 1997-09-19
;   PRIOR APPLICATION NUMBER: 60/059836
;   PRIOR FILING DATE: 1997-09-24
;   PRIOR APPLICATION NUMBER: 60/062250
;   PRIOR FILING DATE: 1997-10-17
;   PRIOR APPLICATION NUMBER: 60/062285
;   PRIOR FILING DATE: 1997-10-17
;   PRIOR APPLICATION NUMBER: 60/062287
;   PRIOR FILING DATE: 1997-10-17
;   PRIOR APPLICATION NUMBER: 60/062814
;   PRIOR FILING DATE: 1997-10-24
;   PRIOR APPLICATION NUMBER: 60/062816
;   PRIOR FILING DATE: 1997-10-24
;   PRIOR APPLICATION NUMBER: 60/063045
;   PRIOR FILING DATE: 1997-10-24
;   PRIOR APPLICATION NUMBER: 60/063082
;   PRIOR FILING DATE: 1997-10-31
;   PRIOR APPLICATION NUMBER: 60/063127
;   PRIOR FILING DATE: 1997-10-24
;   PRIOR APPLICATION NUMBER: 60/063327
;   PRIOR FILING DATE: 1997-10-27
;   PRIOR APPLICATION NUMBER: 60/063329
;   PRIOR FILING DATE: 1997-10-27
;   PRIOR APPLICATION NUMBER: 60/063550
;   PRIOR FILING DATE: 1997-10-28
;   PRIOR APPLICATION NUMBER: 60/063561
;   PRIOR FILING DATE: 1997-10-28
;   PRIOR APPLICATION NUMBER: 60/063704
;   PRIOR FILING DATE: 1997-10-29
;   PRIOR APPLICATION NUMBER: 60/063733
;   PRIOR FILING DATE: 1997-10-29
;   PRIOR APPLICATION NUMBER: 60/063735
;   PRIOR FILING DATE: 1997-10-29
;   PRIOR APPLICATION NUMBER: 60/063738
;   PRIOR FILING DATE: 1997-10-29
;   PRIOR APPLICATION NUMBER: 60/063755
;   PRIOR FILING DATE: 1997-10-17
;   PRIOR APPLICATION NUMBER: 60/064248
;   PRIOR FILING DATE: 1997-11-03
;   PRIOR APPLICATION NUMBER: 60/064809

PRIOR FILING DATE: 1997-11-07
PRIOR APPLICATION NUMBER: 60/065186
PRIOR FILING DATE: 1997-11-12
PRIOR APPLICATION NUMBER: 60/065846
PRIOR FILING DATE: 1997-11-17
PRIOR APPLICATION NUMBER: 60/066364
PRIOR FILING DATE: 1997-11-21
PRIOR APPLICATION NUMBER: 60/066453
PRIOR FILING DATE: 1997-11-24
PRIOR APPLICATION NUMBER: 60/066511
PRIOR FILING DATE: 1997-11-24
PRIOR APPLICATION NUMBER: 60/066770
PRIOR FILING DATE: 1997-11-24
PRIOR APPLICATION NUMBER: 60/069212
PRIOR FILING DATE: 1997-12-11
PRIOR APPLICATION NUMBER: 60/069278
PRIOR FILING DATE: 1997-12-11
PRIOR APPLICATION NUMBER: 60/069334
PRIOR FILING DATE: 1997-12-11
PRIOR APPLICATION NUMBER: 60/069694
PRIOR FILING DATE: 1997-12-16
PRIOR APPLICATION NUMBER: 60/072320
PRIOR FILING DATE: 1998-01-23
PRIOR APPLICATION NUMBER: 60/073612
PRIOR FILING DATE: 1998-02-04
PRIOR APPLICATION NUMBER: 60/074086
PRIOR FILING DATE: 1998-02-09
PRIOR APPLICATION NUMBER: 60/074092
PRIOR FILING DATE: 1998-02-09
PRIOR APPLICATION NUMBER: 60/079728
PRIOR FILING DATE: 1998-02-27
PRIOR APPLICATION NUMBER: 60/079663
PRIOR FILING DATE: 1998-03-25
PRIOR APPLICATION NUMBER: 60/079294
PRIOR FILING DATE: 1998-03-20
PRIOR APPLICATION NUMBER: 60/078910
PRIOR FILING DATE: 1998-03-12
PRIOR APPLICATION NUMBER: 60/077791
PRIOR FILING DATE: 1998-03-12
PRIOR APPLICATION NUMBER: 60/080165
PRIOR FILING DATE: 1998-03-31
PRIOR APPLICATION NUMBER: 60/081203
PRIOR FILING DATE: 1998-04-09
PRIOR APPLICATION NUMBER: 60/081229
PRIOR FILING DATE: 1998-04-09
PRIOR APPLICATION NUMBER: 60/081695
PRIOR FILING DATE: 1998-04-14
PRIOR APPLICATION NUMBER: 60/081817
PRIOR FILING DATE: 1998-04-15
PRIOR APPLICATION NUMBER: 60/081818
PRIOR FILING DATE: 1998-04-15
PRIOR APPLICATION NUMBER: 60/082999
PRIOR FILING DATE: 1998-04-24
PRIOR APPLICATION NUMBER: 60/083322
PRIOR FILING DATE: 1998-04-28
PRIOR APPLICATION NUMBER: 60/083545
PRIOR FILING DATE: 1998-04-29
PRIOR APPLICATION NUMBER: 60/084600
PRIOR FILING DATE: 1998-05-07
PRIOR APPLICATION NUMBER: 60/084627
PRIOR FILING DATE: 1998-05-07
PRIOR APPLICATION NUMBER: 60/084637
PRIOR FILING DATE: 1998-05-07
PRIOR APPLICATION NUMBER: 60/085149
PRIOR FILING DATE: 1998-05-12
PRIOR APPLICATION NUMBER: 60/085323
PRIOR FILING DATE: 1998-05-13
PRIOR APPLICATION NUMBER: 60/085338
PRIOR FILING DATE: 1998-05-13
PRIOR APPLICATION NUMBER: 60/085339
PRIOR FILING DATE: 1998-05-15
PRIOR APPLICATION NUMBER: 60/085579
PRIOR APPLICATION NUMBER: 60/085697
PRIOR FILING DATE: 1999-05-15
PRIOR APPLICATION NUMBER: 60/085704
PRIOR FILING DATE: 1998-05-15
PRIOR APPLICATION NUMBER: 60/086414
PRIOR FILING DATE: 1998-05-22
PRIOR APPLICATION NUMBER: 60/086430
PRIOR FILING DATE: 1998-05-22
PRIOR APPLICATION NUMBER: 60/087106
PRIOR FILING DATE: 1998-05-28
PRIOR APPLICATION NUMBER: 60/087730
PRIOR FILING DATE: 1998-06-04
PRIOR APPLICATION NUMBER: 60/088026
PRIOR FILING DATE: 1998-06-10
PRIOR APPLICATION NUMBER: 60/088741
PRIOR FILING DATE: 1998-06-10
PRIOR APPLICATION NUMBER: 60/088810
PRIOR FILING DATE: 1998-06-10
PRIOR APPLICATION NUMBER: 60/088858
PRIOR FILING DATE: 1998-06-11
PRIOR APPLICATION NUMBER: 60/089532
PRIOR FILING DATE: 19/98-06-11
PRIOR APPLICATION NUMBER: 60/089599
PRIOR FILING DATE: 1998-06-17
PRIOR APPLICATION NUMBER: 60/089907
PRIOR FILING DATE: 1998-06-17
PRIOR APPLICATION NUMBER: 60/089947
PRIOR FILING DATE: 1998-06-18
PRIOR APPLICATION NUMBER: 60/090349
PRIOR FILING DATE: 1998-06-19
PRIOR APPLICATION NUMBER: 60/090429
PRIOR FILING DATE: 1998-06-23
PRIOR APPLICATION NUMBER: 60/090445
PRIOR FILING DATE: 1998-06-24
PRIOR APPLICATION NUMBER: 60/090538
PRIOR FILING DATE: 1998-06-24
PRIOR APPLICATION NUMBER: 60/090863
PRIOR FILING DATE: 1998-06-24
PRIOR APPLICATION NUMBER: 60/091360
PRIOR FILING DATE: 1998-06-26
PRIOR APPLICATION NUMBER: 60/091519
PRIOR FILING DATE: 1998-07-01
PRIOR APPLICATION NUMBER: 60/091982
PRIOR FILING DATE: 1998-07-02
PRIOR APPLICATION NUMBER: 60/091982
PRIOR FILING DATE: 1998-07-07

```
Query Match      14.5%;    Score 278.5;   DB 9;     Length 765;
Best Local Similarity   25.0%;    Pred. No. 9e-20;    Indels   45;   Gaps  9;
Matches   80;   Conservative   76;   Mismatches   119;

Qy   72  RQLYVASAICLLFMIGEVVGGYLAHSLAVMTDAAHLLTDFASMLISLFSLWMSSRPATKT  131
             |!:   |||  |!: | | ||  | |  :|||  | |:| | :::| |  ||  |  !:
Db  417  RQIFYFLCLNLLFTFVELFVGVLTNSLGLISDGFHHMLFDCSALVMGLFAALMSRWKATRI  476

Qy  132  MNFGWQRAEILGALVSVLSIWVVTGVLVYLLAVERLISGDYEIDGGTMLITSGCAVAVNII  191
             :|:| |||  |||   || | :|| | | :|  | | |:|    |  | | || | | !:|
Db  477  FSYGYGRIEILSGFINGLFLIVIAFVFVMESVARLIDPP-ELDTHMLITPVSVGGLIVNLI  535

Qy  192  MGL---------------------TLHQS--------------GHGHSHGTTNQQEENPSVRAAF  221
             |:|                     | |              ||||||  |   |:|
Db  536  -GICAFSHAHSHAHGASGQGSCHSSDHSHSHHMHGHSDHGHGHSHSGSAG-GGMNANMRGVF  593

Qy  222  IHVIGDFMQSMGVLVAAYILYFKPEYKYVDPICTFVFSILVLGTTLTLRDVILLVLMEGT  281
             :|!: | :||||!: | | :|: | :|  | !: |  |  :|  !:
Db  594  LHVLADTLGSLGVIVST-VLIEQFGWFIADPLCSLSTAILIFLSVVPLIKDACQVLLRL  652

Qy  282  PKGVDFTAVRDLLLSVEGVEALHSL------HIWALTVAQPVLSVHIAIAQNTDAQAVLK  335
             | : |::|| : | |::|   |!:|     :|| | |  ::  :|| :  :
Db  653  PPEYE----KELHIALEKIQKIEGLISYRDPHFWRHSASIVAGTIHIQVTSDVLEQRIVQ  708

Qy  336  TASSRLQGKFHFHTVTIQIE  355
             :| :  |: :|!:|
Db  709  QVTGILKDA-GVNNLIQVE  727
```

RESULT 12
US-10-121-049-28
; Sequence 28, Application US/10121049
; Publication No. US20030022239A1
; GENERAL INFORMATION:
APPLICANT: Beresini,Maureen
APPLICANT: DeForge,Laura
APPLICANT: Desnoyers,Luc
APPLICANT: Filvaroff,Ellen
APPLICANT: Gao,Wei-Qiang
APPLICANT: Gerritsen,Mary E.
APPLICANT: Goddard,Audrey
APPLICANT: Godowski,Paul J.
APPLICANT: Gurney,Austin L.
APPLICANT: Sherwood,Steven
APPLICANT: Smith,Victoria
APPLICANT: Stewart,Timothy A.
APPLICANT: Tumas,Daniel
APPLICANT: Watanabe,Colin K
APPLICANT: Wood,William
APPLICANT: Zhang, Zemin
TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
TITLE OF INVENTION: ACIDS ENCODING THE SAME
FILE REFERENCE: P3330R1C17
CURRENT APPLICATION NUMBER: US/10/121,049
CURRENT FILING DATE:
  Prior Application removed - See File Wrapper or Palm
NUMBER OF SEQ ID NOS: 550
SEQ ID NO 28
  LENGTH: 765
  TYPE: PRT
  ORGANISM: Homo Sapien
US-10-121-049-28

Query Match   14.5%; Score 278.5; DB 9; Length 765;
Best Local Similarity   25.0%; Pred. No. 9e-20;
Matches   80; Conservative   76; Mismatches 119; Indels   45; Gaps   9;

```
Qy  72 RQLYVASAICLLFMIGEVVGGYLAHSLAVMTDAAHLLTDFASMLISLFSLKWMSSRPATKT 131
        |:  |||   : ||| ||    ::|    ||:|   ||::
Db 417 RQIFYFLCLNLLFTVELFYGVLTNSLGLISDGFHMLFDCSALVMGLFAALMSRWKATRI 476

Qy 132 MNFGWQRAEILGALVSVLSIWVVTGVLVLVYLAVERLISGDYEIDGGTMLITSGCAVAVNII 191
        |  ::| |: ||  |: |:    ::| ||| ||   |:| |     :|||
Db 477 FSYGYGRIEILSGFINGLFLIVIAFFVFMESVARLIDPP-ELDTHMLTPVSVGGLIVNLI 535

Qy 192 MGL------------------------TLHQS------GHGHSHGTTNQQEENPSVRAAF 221
        |                          :| |       ||||||| :
Db 536 -GICAFSHAHSHAHGASQGSCHSSDSHSHSHMHGHSDHGHGHSHGSSAG-GGMNANMRGVF 593

Qy 222 IHVIGDFMQSMGVLVAAYILYFKPEYKYVDPICTEVFSILVLGTTLTILRDVILVLMEGT 281
        :||: ||  :| |:| :  |  || :||:|   |   || |:  || :  || :
Db 594 LHVLADTLGSIGVIVST-VLIEQFGWFIADPLCSLSTAILIFLSVVPLIKDACQVLLLRL 652

Qy 282 PKGVDFTAVRDLLLSVEGVEALHSL------HIWALTVAQPVLLSVHIAIAQNTDAQAVLK 335
        | : |::|  |: :|::|  ||:|       ::|  |  |||:|| ::|| ::
Db 653 PPEYE---KELHIALEKIQKEGLISYRDPHFWRHSASIVAGTIHIQVTSDVLEQRIVQ 708

Qy 336 TASSRLQGKFHFHTVTIQIE 355
        |:  | : ::|||: |
Db 709 QVTGILKDA-GVNNLTIQVE 727
```

RESULT 13
US-10-123-904-28
; Sequence 28, Application US/10123904
; Publication No. US20030022328A1
; GENERAL INFORMATION:
APPLICANT: Baker,Kevin P.
APPLICANT: Beresini,Maureen
APPLICANT: DeForge,Laura
APPLICANT: Desnoyers,Luc
APPLICANT: Filvaroff,Ellen
APPLICANT: Gao,Wei-Qiang
APPLICANT: Gerritsen,Mary E.
APPLICANT: Goddard,Audrey
APPLICANT: Godowski,Paul J.
APPLICANT: Gurney,Austin L.
APPLICANT: Sherwood,Steven
APPLICANT: Smith,Victoria

APPLICANT: Desnoyers,Luc
APPLICANT: Filvaroff,Ellen
APPLICANT: Gao,Wei-Qiang
APPLICANT: Gerritsen,Mary E.
APPLICANT: Goddard,Audrey
APPLICANT: Godowski,Paul J.
APPLICANT: Gurney,Austin L.
APPLICANT: Sherwood,Steven
APPLICANT: Smith,Victoria
APPLICANT: Stewart,Timothy A.
APPLICANT: Tumas,Daniel
APPLICANT: Watanabe,Colin K
APPLICANT: Wood,William
APPLICANT: Zhang, Zemin
TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
TITLE OF INVENTION: ACIDS ENCODING THE SAME
FILE REFERENCE: P3330R1C54
CURRENT APPLICATION NUMBER: US/10/123,904
CURRENT FILING DATE: 2002-04-16
  Prior Application removed - See File Wrapper or Palm
NUMBER OF SEQ ID NOS: 550
SEQ ID NO 28
  LENGTH: 765
  TYPE: PRT
  ORGANISM: Homo Sapien
US-10-123-904-28

Query Match   14.5%; Score 278.5; DB 9; Length 765;
Best Local Similarity   25.0%; Pred. No. 9e-20;
Matches   80; Conservative   76; Mismatches 119; Indels   45; Gaps   9;

```
Qy  72 RQLYVASAICLLFMIGEVVGGYLAHSLAVMTDAAHLLTDFASMLISLFSLKWMSSRPATKT 131
        |:  |||   : ||| ||    ::|    ||:|   ||::
Db 417 RQIFYFLCLNLLFTVELFYGVLTNSLGLISDGFHMLFDCSALVMGLFAALMSRWKATRI 476

Qy 132 MNFGWQRAEILGALVSVLSIWVVTGVLVLVYLAVERLISGDYEIDGGTMLITSGCAVAVNII 191
        |  ::| |: ||  |: |:    ::| ||| ||   |:| |     :|||
Db 477 FSYGYGRIEILSGFINGLFLIVIAFFVFMESVARLIDPP-ELDTHMLTPVSVGGLIVNLI 535

Qy 192 MGL------------------------TLHQS------GHGHSHGTTNQQEENPSVRAAF 221
        |                          :| |       ||||||| :
Db 536 -GICAFSHAHSHAHGASQGSCHSSDSHSHSHMHGHSDHGHGHSHGSSAG-GGMNANMRGVF 593

Qy 222 IHVIGDFMQSMGVLVAAYILYFKPEYKYVDPICTEVFSILVLGTTLTILRDVILVLMEGT 281
        :||: ||  :| |:| :  |  || :||:|   |   || |:  || :  || :
Db 594 LHVLADTLGSIGVIVST-VLIEQFGWFIADPLCSLSTAILIFLSVVPLIKDACQVLLLRL 652

Qy 282 PKGVDFTAVRDLLLSVEGVEALHSL------HIWALTVAQPVLLSVHIAIAQNTDAQAVLK 335
        | : |::|  |: :|::|  ||:|       ::|  |  |||:|| ::|| ::
Db 653 PPEYE---KELHIALEKIQKEGLISYRDPHFWRHSASIVAGTIHIQVTSDVLEQRIVQ 708

Qy 336 TASSRLQGKFHFHTVTIQIE 355
        |:  | : ::|||: |
Db 709 QVTGILKDA-GVNNLTIQVE 727
```

RESULT 14
US-10-140-470-28
; Sequence 28, Application US/10140470
; Publication No. US20030022331A1
; GENERAL INFORMATION:
APPLICANT: Baker,Kevin P.
APPLICANT: Beresini,Maureen
APPLICANT: DeForge,Laura
APPLICANT: Desnoyers,Luc
APPLICANT: Filvaroff,Ellen
APPLICANT: Gao,Wei-Qiang
APPLICANT: Gerritsen,Mary E.
APPLICANT: Goddard,Audrey
APPLICANT: Godowski,Paul J.
APPLICANT: Gurney,Austin L.
APPLICANT: Sherwood,Steven
APPLICANT: Smith,Victoria

APPLICANT: Stewart,Timothy A.
APPLICANT: Tumas,Daniel
APPLICANT: Watanabe,Colin K
APPLICANT: Wood,William
APPLICANT: Zhang, Zemin
TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
TITLE OF INVENTION: ACIDS ENCODING THE SAME
FILE REFERENCE: P3330R1C160
CURRENT APPLICATION NUMBER: US/10/140,470
CURRENT FILING DATE: 2002-05-06
Prior Application removed - See Palm or File Wrapper
NUMBER OF SEQ ID NOS: 550
SEQ ID NO 28
  TYPE: PRT
  ORGANISM: Homo Sapien
US-10-140-470-28

Query Match   14.5%; Score 278.5; DB 9; Length 765;
Best Local Similarity   25.0%; Pred. No. 9e-20;
Matches   80; Conservative   76; Mismatches 119; Indels   45; Gaps   9;

Qy   72 RQLYVASAICLLFMIGEVVGGYLAHSLAVWTDAAHLLTDFASMLISLFSLWMSSRPATKT 131
          ||:| ||| ||: |: |:| || :|:| :.|:.||: || ||||:|| | :
Db 417 RQIFYFLCLNLLFTFVELFYGVLTNSLGLISDGFHMLFDCSALVMGLFAALMSRWKATRI 476

Qy 132 MNFGWQRAEILGALVSVLSIWVVTGVLVLYLAVERLISGDYEIDGGTMLITSGGCAVAVNII 191
          :|:| |||: || || | || || :||| || : : | |:::| |:||||:|
Db 477 FSYGYGRIEILSGFINGLFLIVIAFFVFMESVARLIDPP--ELDTHMLTPVSVGGLIVNLI 535

Qy 192 MGL--------------TLHQS------------GHGHSHGTTNQQEENPSVRAAF 221
          :| | | |||||| : : : || |
Db 536 -GICAFSHAHSHAHGASQGSCHSSDHSHSHHMGHSDHGHGHSHGSAG-GGMNANMRGVF 593

Qy 222 IHVIGDEFMQSMGVLVAAYILYFKPEYKYVDPICTFVFSILVLGTTLTILRDVILVLMEGT 281
          :|| | : |:| | || :| | :| : |:| :::| || ||:|||||:|:|:|
Db 594 LHVLADTLGSIGVIVST-VLIEQFGWFIADPLCSLSTAILIFLSVVPLIKDACQVLLRL 652

Qy 282 PKGVDFTAVRDLLLSVEGVEALHSL------ -HIWALTVAQPVLSVHIAIAQNTDAQAVLK 335
          | :   : |: ::| :| :::   :| || | :|::| :|:: ::: | :|
Db 653 PPEYE---KELHIALEKIQKIEGLISYRDPHFWRHSASIVAGTIHIQVTSDVLEQRIVQ 708

Qy 336 TASSRLQGKFHFHTVTIQIE 355
          || |::|||:|| |||:|
Db 709 QVTGILKDA-GVNNLTIQVE 727

RESULT 15
US-10-175-746-28
Sequence 28, Application US/10175746
Publication No. US20030027270A1
GENERAL INFORMATION:
  APPLICANT: Baker,Kevin P.
  APPLICANT: Beresini,Maureen
  APPLICANT: Deforge,Laura
  APPLICANT: Desnoyers,Luc
  APPLICANT: Filvaroff,Ellen
  APPLICANT: Gao,Wei-Qiang
  APPLICANT: Gerritsen,Mary E.
  APPLICANT: Goddard,Audrey
  APPLICANT: Godowski,Paul J.
  APPLICANT: Gurney,Austin L.
  APPLICANT: Sherwood,Steven
  APPLICANT: Smith,Victoria
  APPLICANT: Stewart,Timothy A.
  APPLICANT: Tumas,Daniel
  APPLICANT: Watanabe,Colin K
  APPLICANT: Wood,William
  APPLICANT: Zhang, Zemin
  TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
  TITLE OF INVENTION: ACIDS ENCODING THE SAME
  FILE REFERENCE: P3330R1C353
  CURRENT APPLICATION NUMBER: US/10/175,746

CURRENT FILING DATE: 2002-06-19
Prior Application removed - See File Wrapper or Palm
NUMBER OF SEQ ID NOS: 550
SEQ ID NO 28
  TYPE: PRT
  ORGANISM: Homo Sapien
US-10-175-746-28

Query Match   14.5%; Score 278.5; DB 9; Length 765;
Best Local Similarity   25.0%; Pred. No. 9e-20;
Matches   80; Conservative   76; Mismatches 119; Indels   45; Gaps   9;

Qy   72 RQLYVASAICLLFMIGEVVGGYLAHSLAVWTDAAHLLTDFASMLISLFSLWMSSRPATKT 131
          ||:| ||| ||: |: |:| || :|:| :.|:.||: || ||||:|| | :
Db 417 RQIFYFLCLNLLFTFVELFYGVLTNSLGLISDGFHMLFDCSALVMGLFAALMSRWKATRI 476

Qy 132 MNFGWQRAEILGALVSVLSIWVVTGVLVLYLAVERLISGDYEIDGGTMLITSGGCAVAVNII 191
          :|:| |||: || || | || || :||| || : : | |:::| |:||||:|
Db 477 FSYGYGRIEILSGFINGLFLIVIAFFVFMESVARLIDPP--ELDTHMLTPVSVGGLIVNLI 535

Qy 192 MGL--------------TLHQS------------GHGHSHGTTNQQEENPSVRAAF 221
          :| | | |||||| : : : || |
Db 536 -GICAFSHAHSHAHGASQGSCHSSDHSHSHHMGHSDHGHGHSHGSAG-GGMNANMRGVF 593

Qy 222 IHVIGDEFMQSMGVLVAAYILYFKPEYKYVDPICTFVFSILVLGTTLTILRDVILVLMEGT 281
          :|| | : |:| | || :| | :| : |:| :::| || ||:|||||:|:|:|
Db 594 LHVLADTLGSIGVIVST-VLIEQFGWFIADPLCSLSTAILIFLSVVPLIKDACQVLLRL 652

Qy 282 PKGVDFTAVRDLLLSVEGVEALHSL------ -HIWALTVAQPVLSVHIAIAQNTDAQAVLK 335
          | :   : |: ::| :| :::   :| || | :|::| :|:: ::: | :|
Db 653 PPEYE---KELHIALEKIQKIEGLISYRDPHFWRHSASIVAGTIHIQVTSDVLEQRIVQ 708

Qy 336 TASSRLQGKFHFHTVTIQIE 355
          || |::|||:|| |||:|
Db 709 QVTGILKDA-GVNNLTIQVE 727

RESULT 16
US-10-176-918-28
Sequence 28, Application US/10176918
Publication No. US20030027275A1
GENERAL INFORMATION:
  APPLICANT: Baker,Kevin P.
  APPLICANT: Beresini,Maureen
  APPLICANT: Deforge,Laura
  APPLICANT: Desnoyers,Luc
  APPLICANT: Filvaroff,Ellen
  APPLICANT: Gao,Wei-Qiang
  APPLICANT: Gerritsen,Mary E.
  APPLICANT: Goddard,Audrey
  APPLICANT: Godowski,Paul J.
  APPLICANT: Gurney,Austin L.
  APPLICANT: Sherwood,Steven
  APPLICANT: Smith,Victoria
  APPLICANT: Stewart,Timothy A.
  APPLICANT: Tumas,Daniel
  APPLICANT: Watanabe,Colin K
  APPLICANT: Wood,William
  APPLICANT: Zhang, Zemin
  TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
  TITLE OF INVENTION: ACIDS ENCODING THE SAME
  FILE REFERENCE: P3330R1C382
  CURRENT APPLICATION NUMBER: US/10/176,918
  CURRENT FILING DATE: 2002-06-20
  Prior Application removed - See File Wrapper or Palm
  NUMBER OF SEQ ID NOS: 550
SEQ ID NO 28
  TYPE: PRT
  ORGANISM: Homo Sapien
US-10-176-918-28

Qy   Query Match   14.5%;   Score 278.5;   DB 9;   Length 765;
      Best Local Similarity 25.0%;   Pred. No. 9e-20;
      Matches   80;   Conservative   76;   Mismatches 119;   Indels   45;   Gaps   9;

Qy   72 RQLYVASAICLLFMIGEVVGGYLAHSLAVWTDAAHLLTDFASMLISLFSLWMSSRPATKT 131
           |:||:| || |:|| ::|:| | | |:| |||:|:|| || || ::|
Db   417 RQIFYFLCLNLLFTFVELFYGVLTNSLGLISDGFHMLFDCSALVMGLFAALMSRWKATRI 476

Qy   132 MNFGWQRAEILGALVSVLSIWVTGVLVYLAVERLISGDYEIDGGTMLITSGCAVAVNII 191
           ::|| ||||:|||||| ||| |:|| | |:|:| | |:| :: | : | | ||
Db   477 FSYGYGRIEILSGFINGLFLIVIAFFVFMESVARLIDPP-ELDTHMLTPVSVGGLIVNLI 535

Qy   192 MGL------------------------------TLHQS------------------------------ 221
           |:| | |
Db   536 -GICAFSHAHSHAHGASQGSCHSSDHSHSHHMHGHSDHGHGHSHSGSAG-GGMNANMRGVF 593

Qy   222 IHVIGDFMQSMGVLVAAYILYFKPEYKYVDPICTFVFSILVLGTTLTILRDVILVLMEGT 281
           :||:| | |:|: | |: | ||| ||: :| :: |:| ||: ::|:: |:|:| ||:
Db   594 LHVLADTLGSIGVIVST-VLIEQFGWFIADPLCSLSTAILIFLSVVPLIKDACQVLLLRL 652

Qy   282 PKGVDFTAVRDLLLSVEGVEALHSL------HIWALTVAQPVLSVHHIAIAQNTDAQAVLK 335
           | | | | :|||:::| | |||| || | ||: || || |
Db   653 PPEYE---KELHIALEKIQKIEGLISYRDPHFWRHSASIVAGTIHHQVTSDVLEQRIVQ 708

Qy   336 TASSRLQGKFHFHTVTIQIE 355
           |: ||||| | | |:|
Db   709 QVTGILKDA-GVNNLTIQVE 727

RESULT 17
US-10-176-921-28
Sequence 28, Application US/10176921
Publication No. US20030027276A1
GENERAL INFORMATION:
 ; APPLICANT: Baker,Kevin P.
 ; APPLICANT: Beresini,Maureen
 ; APPLICANT: DeForge,Laura
 ; APPLICANT: Desnoyers,Luc
 ; APPLICANT: Filvaroff,Ellen
 ; APPLICANT: Gao,Wei-Qiang
 ; APPLICANT: Gerritsen,Mary E.
 ; APPLICANT: Goddard,Audrey
 ; APPLICANT: Godowski,Paul J.
 ; APPLICANT: Gurney,Austin L.
 ; APPLICANT: Sherwood,Steven
 ; APPLICANT: Smith,Victoria
 ; APPLICANT: Stewart,Timothy A.
 ; APPLICANT: Tumas,Daniel
 ; APPLICANT: Watanabe,Colin K
 ; APPLICANT: Wood,William
 ; APPLICANT: Zhang, Zemin
 ; TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
 ; TITLE OF INVENTION: ACIDS ENCODING THE SAME
 ; FILE REFERENCE: P3330R1C288
 ; CURRENT APPLICATION NUMBER: US/10/176,921
 ; CURRENT FILING DATE: 2002-06-20
 ; PRIOR APPLICATION NUMBER: US/10/176921
 ; PRIOR Application removed - See File Wrapper or Palm
 ; NUMBER OF SEQ ID NOS: 550
 ; SEQ ID NO 28
 ; LENGTH: 765
 ; TYPE: PRT
 ; ORGANISM: Homo Sapien
US-10-176-921-28

Qy   Query Match   14.5%;   Score 278.5;   DB 9;   Length 765;
      Best Local Similarity 25.0%;   Pred. No. 9e-20;
      Matches   80;   Conservative   76;   Mismatches 119;   Indels   45;   Gaps   9;

Qy   72 RQLYVASAICLLFMIGEVVGGYLAHSLAVWTDAAHLLTDFASMLISLFSLWMSSRPATKT 131
           |:||:| || |:|| ::|:| | | |:| |||:|:|| || || ::|
Db   417 RQIFYFLCLNLLFTFVELFYGVLTNSLGLISDGFHMLFDCSALVMGLFAALMSRWKATRI 476

Qy   132 MNFGWQRAEILGALVSVLSIWVTGVLVYLAVERLISGDYEIDGGTMLITSGCAVAVNII 191

Db   477 FSYGYGRIEILSGFINGLFLIVIAFFVFMESVARLIDPP-ELDTHMLTPVSVGGLIVNLI 535

Qy   192 MGL------------------------------TLHQS------------------------------ 221
           |:| | |
Db   536 -GICAFSHAHSHAHGASQGSCHSSDHSHSHHMHGHSDHGHGHSHSGSAG-GGMNANMRGVF 593

Qy   222 IHVIGDFMQSMGVLVAAYILYFKPEYKYVDPICTFVFSILVLGTTLTILRDVILVLMEGT 281
           :||:| | |:|: | |: | ||| ||: :| :: |:| ||: ::|:: |:|:| ||:
Db   594 LHVLADTLGSIGVIVST-VLIEQFGWFIADPLCSLSTAILIFLSVVPLIKDACQVLLLRL 652

Qy   282 PKGVDFTAVRDLLLSVEGVEALHSL------HIWALTVAQPVLSVHHIAIAQNTDAQAVLK 335
Db   653 PPEYE---KELHIALEKIQKIEGLISYRDPHFWRHSASIVAGTIHHQVTSDVLEQRIVQ 708

Qy   336 TASSRLQGKFHFHTVTIQIE 355
Db   709 QVTGILKDA-GVNNLTIQVE 727

RESULT 18
US-10-137-865-28
Sequence 28, Application US/10137865
Publication No. US20030032155A1
GENERAL INFORMATION:
 ; APPLICANT: Baker,Kevin P.
 ; APPLICANT: Beresini,Maureen
 ; APPLICANT: DeForge,Laura
 ; APPLICANT: Desnoyers,Luc
 ; APPLICANT: Filvaroff,Ellen
 ; APPLICANT: Gao,Wei-Qiang
 ; APPLICANT: Gerritsen,Mary E.
 ; APPLICANT: Goddard,Audrey
 ; APPLICANT: Godowski,Paul J.
 ; APPLICANT: Gurney,Austin L.
 ; APPLICANT: Sherwood,Steven
 ; APPLICANT: Smith,Victoria
 ; APPLICANT: Stewart,Timothy A.
 ; APPLICANT: Tumas,Daniel
 ; APPLICANT: Watanabe,Colin K
 ; APPLICANT: Wood,William
 ; APPLICANT: Zhang, Zemin
 ; TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
 ; TITLE OF INVENTION: ACIDS ENCODING THE SAME
 ; FILE REFERENCE: P3330R1C154
 ; CURRENT APPLICATION NUMBER: US/10/137,865
 ; CURRENT FILING DATE: 2002-05-03
 ; PRIOR Application removed - See Palm or File Wrapper
 ; NUMBER OF SEQ ID NOS: 550
 ; SEQ ID NO 28
 ; LENGTH: 765
 ; TYPE: PRT
 ; ORGANISM: Homo Sapien
US-10-137-865-28

Qy   Query Match   14.5%;   Score 278.5;   DB 9;   Length 765;
      Best Local Similarity 25.0%;   Pred. No. 9e-20;
      Matches   80;   Conservative   76;   Mismatches 119;   Indels   45;   Gaps   9;

Qy   72 RQLYVASAICLLFMIGEVVGGYLAHSLAVWTDAAHLLTDFASMLISLFSLWMSSRPATKT 131
           |:||:| || |:|| ::|:| | | |:| |||:|:|| || || ::|
Db   417 RQIFYFLCLNLLFTFVELFYGVLTNSLGLISDGFHMLFDCSALVMGLFAALMSRWKATRI 476

Qy   132 MNFGWQRAEILGALVSVLSIWVTGVLVYLAVERLISGDYEIDGGTMLITSGCAVAVNII 191

Db   477 FSYGYGRIEILSGFINGLFLIVIAFFVFMESVARLIDPP-ELDTHMLTPVSVGGLIVNLI 535

Qy   192 MGL------------------------------TLHQS------------------------------ 221
Db   536 -GICAFSHAHSHAHGASQGSCHSSDHSHSHHMHGHSDHGHGHSHSGSAG-GGMNANMRGVF 593

Qy   222 IHVIGDFMQSMGVLVAAYILYFKPEYKYVDPICTFVFSILVLGTTLTILRDVILVLMEGT 281
Db   594

```
Db  594  LHVLADTLGSIGVIVST-VLIEQFGWFIADPLCSLSTAILIFLSVVPLIKDACQVLLLRL 652
                                                                            
Qy  282  PKGVDFTAVRDLLLSVEGVEALHSL------HIWALTVAQPVLSVHIAIAQNTDAQAVLK 335
                                                                            
Db  653  PPEYE---KELHIALEKIQKIEGLISYRDPHFWRHSASIVAGTIHIQVTSDVLEQRIVQ 708
                                                                            
Qy  336  TASSRLQGKFHFHTVTIQIE 355
                                                                            
Db  709  QVTGILKDA-GVNNLTIQVE 727
```

RESULT 19
US-10-140-474-28
Sequence 28, Application US/10140474
Publication No. US20030032156A1
GENERAL INFORMATION:
APPLICANT: Baker, Kevin P.
APPLICANT: Beresini, Maureen
APPLICANT: DeForge, Laura
APPLICANT: Desnoyers, Luc
APPLICANT: Filvaroff, Ellen
APPLICANT: Gao, Wei-Qiang
APPLICANT: Gerritsen, Mary E.
APPLICANT: Goddard, Audrey
APPLICANT: Godowski, Paul J.
APPLICANT: Gurney, Austin L.
APPLICANT: Sherwood, Steven
APPLICANT: Smith, Victoria
APPLICANT: Stewart, Timothy A.
APPLICANT: Tumas, Daniel
APPLICANT: Watanabe, Colin K
APPLICANT: Wood, William
APPLICANT: Zhang, Zemin
TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
TITLE OF INVENTION: ACIDS ENCODING THE SAME
FILE REFERENCE: P3330R1C162
CURRENT APPLICATION NUMBER:
CURRENT FILING DATE: 2002-05-06
Prior Application removed - See Palm or File Wrapper
NUMBER OF SEQ ID NOS: 550
SEQ ID NO 28
LENGTH: 765
TYPE: PRT
ORGANISM: Homo Sapien
US-10-140-474-28

Query Match   14.5%;   Score 278.5;   DB 9;   Length 765;
Best Local Similarity   25.0%;   Pred. No. 9e-20;
Matches   80;   Conservative   76;   Mismatches   119;   Indels   45;   Gaps   9;

```
Qy   72  ROLVASAICLLFMIGEVVGGYLAHSLAVMTDAAHLLTDFASMLISLFSLWMSSRPATKT 131
Db  417  ROIFYLCLNLLFTVELFYGVLTNSLGLISDGFHMLFDCSALVMGLFAALMSRWKATRI 476

Qy  132  MNFGWQRAEILGALVSVLSIWVTGVLVYLAVERLISGDYEIDGGTMLITSGCAVAVNII 191
Db  477  FSYGYGRIEILSGFINGLFLIVIAFFVEMESVARLIDPP-ELDTHMLTPVSVGGLIVNLI 535

Qy  192  MGL----------------------TLHQS------------GHGHSHGTTNQQEENPSVRAAF 221
Db  536  -GICAFSHAHSHAHGASQGSCHSSDHSHSHHMHGHSDHGHGHSHGSAG-GGMNANMRGVF 593

Qy  222  IHVIGDFMQSMGVLVAAYILYFKPEYKYVDPICTFVFSILVLGTTLTILRDVILVLMEGT 281
Db  594  LHVLADTLGSIGVIVST-VLIEQFGWFIADPLCSLSTAILIFLSVVPLIKDACQVLLLRL 652

Qy  282  PKGVDFTAVRDLLLSVEGVEALHSL------HIWALTVAQPVLSVHIAIAQNTDAQAVLK 335
Db  653  PPEYE---KELHIALEKIQKIEGLISYRDPHFWRHSASIVAGTIHIQVTSDVLEQRIVQ 708

Qy  336  TASSRLQGKFHFHTVTIQIE 355
Db  709  QVTGILKDA-GVNNLTIQVE 727
```

RESULT 20
US-10-142-431-28
Sequence 28, Application US/10142431
Publication No. US20030036179A1
GENERAL INFORMATION:
APPLICANT: Baker, Kevin P.
APPLICANT: Beresini, Maureen
APPLICANT: DeForge, Laura
APPLICANT: Desnoyers, Luc
APPLICANT: Filvaroff, Ellen
APPLICANT: Gao, Wei-Qiang
APPLICANT: Gerritsen, Mary E.
APPLICANT: Goddard, Audrey
APPLICANT: Godowski, Paul J.
APPLICANT: Gurney, Austin L.
APPLICANT: Sherwood, Steven
APPLICANT: Smith, Victoria
APPLICANT: Stewart, Timothy A.
APPLICANT: Tumas, Daniel
APPLICANT: Watanabe, Colin K
APPLICANT: Wood, William
APPLICANT: Zhang, Zemin
TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
TITLE OF INVENTION: ACIDS ENCODING THE SAME
FILE REFERENCE: P3330R1C251
CURRENT APPLICATION NUMBER: US/10/142,431
CURRENT FILING DATE: 2002-05-10
Prior Application removed - See File Wrapper or Palm
NUMBER OF SEQ ID NOS: 550
SEQ ID NO 28
LENGTH: 765
TYPE: PRT
ORGANISM: Homo Sapien
US-10-142-431-28

Query Match   14.5%;   Score 278.5;   DB 9;   Length 765;
Best Local Similarity   25.0%;   Pred. No. 9e-20;
Matches   80;   Conservative   76;   Mismatches   119;   Indels   45;   Gaps   9;

```
Qy   72  ROLVASAICLLFMIGEVVGGYLAHSLAVMTDAAHLLTDFASMLISLFSLWMSSRPATKT 131
Db  417  ROIFYLCLNLLFTVELFYGVLTNSLGLISDGFHMLFDCSALVMGLFAALMSRWKATRI 476

Qy  132  MNFGWQRAEILGALVSVLSIWVTGVLVYLAVERLISGDYEIDGGTMLITSGCAVAVNII 191
Db  477  FSYGYGRIEILSGFINGLFLIVIAFFVEMESVARLIDPP-ELDTHMLTPVSVGGLIVNLI 535

Qy  192  MGL----------------------TLHQS------------GHGHSHGTTNQQEENPSVRAAF 221
Db  536  -GICAFSHAHSHAHGASQGSCHSSDHSHSHHMHGHSDHGHGHSHGSAG-GGMNANMRGVF 593

Qy  222  IHVIGDFMQSMGVLVAAYILYFKPEYKYVDPICTFVFSILVLGTTLTILRDVILVLMEGT 281
Db  594  LHVLADTLGSIGVIVST-VLIEQFGWFIADPLCSLSTAILIFLSVVPLIKDACQVLLLRL 652

Qy  282  PKGVDFTAVRDLLLSVEGVEALHSL------HIWALTVAQPVLSVHIAIAQNTDAQAVLK 335
Db  653  PPEYE---KELHIALEKIQKIEGLISYRDPHFWRHSASIVAGTIHIQVTSDVLEQRIVQ 708

Qy  336  TASSRLQGKFHFHTVTIQIE 355
Db  709  QVTGILKDA-GVNNLTIQVE 727
```

RESULT 21
US-10-143-114-28
Sequence 28, Application US/10143114
Publication No. US20030036180A1
GENERAL INFORMATION:
APPLICANT: Baker, Kevin P.
APPLICANT: Beresini, Maureen

CURRENT APPLICATION NUMBER: US/10/142,419
CURRENT FILING DATE: 2002-05-10
Prior Application removed - See File Wrapper or Palm
NUMBER OF SEQ ID NOS: 550
SEQ ID NO 28
LENGTH: 765
TYPE: PRT
ORGANISM: Homo Sapien
US-10-142-419-28

Query Match　　14.5%;　Score 278.5;　DB 9;　Length 765;
Best Local Similarity　　25.0%;　Pred. No. 9e-20;
Matches　80;　Conservative　76;　Mismatches 119;　Indels 45;　Gaps 9;

```
Qy   72 RQLYVASAICLLFMIGEVVGGYLAHSLAVMTDAAHLLTDFASMLISLFSLWMSSRPATKT 131
         |!|: ||| |:| |  :|:|!|! |!|||:||!||  ||:  ||  ||  |!:
Db  417 RQIFYFLCLNLLFTFVELFYGVLTNSLGLISDGFHMLFDCSALVGLFAALMSRWKATRI 476

Qy  132 MNFGWQRAEILGALVSVLSIWVVTGVLVYLYLAVERLISGDYEIDGGTMLITSGCAVAVNII 191
         :!|: ||| |  :|: | |:|  | ||| |||
Db  477 FSYGYGRIEILSGFINGLFLIVIAFFVFMESVARLIDPP-ELDTHMLTPVSVGGLIVNLI 535

Qy  192 MGL---------------------TLHQS----------------------GHGHSHGTTNQOEENPSVRAAF 221
                                                      ||||||||| ::|:|
Db  536 -GICAFSHAHSHAHGASQGSCHSSSDHSHSHHMHGHSDHGHGHSHGSSAG-GGMNANMRGVF 593

Qy  222 IHVIGDFMQSMGVLVAAYILYFKPEYKYVDPICTFVFSILVLGTTLTILRDVILVLMEGT 281
         ||:|| : | |:|  ::||:|!||:||!:  ||:||| ::|!|  |!| :
Db  594 LHVLADTLGSIGVIVST-VLIEQFGWFIADPLCSLSTAILIFLSVVPLIKDACQVLLRL 652

Qy  282 PKGVDFTAVRDLLLSVEGVEALHSL------HIWALTVAQPVLSVHIAIAQNTDAQAVLK 335
             :!:| ||  |::   |:!| ||| ||| :|::
Db  653 PPEYE---KELHIALEKIQKIEGLISYRDPHFWRHSASIVAGTIHIQVTSDVLEQRIVQ 708

Qy  336 TASSRLQGKFHFHTVTIQIE 355
               |: :|||||!||:|
Db  709 QVTGILKDA-GVNNLIIQVE 727
```

RESULT 24
US-10-139-262-8
Sequence 8, Application US/10139262
Patent No. US20020128459A1
GENERAL INFORMATION:
APPLICANT: NAKAMURA, YUSUKE
APPLICANT: TANAKA, TOSHIHIRO
APPLICANT: TSUKADA, SHUICHI
TITLE OF INVENTION: PHYSIOLOGICALLY ACTIVE PROTEIN ORIGINATING IN MAMMALS
FILE REFERENCE:
CURRENT APPLICATION NUMBER: US/10/139,262
CURRENT FILING DATE: 2002-05-07
PRIOR APPLICATION NUMBER: US/09/380,287A
PRIOR FILING DATE: 1999-11-16
PRIOR APPLICATION NUMBER: JP　　1997-62259
PRIOR FILING DATE: 1997-02-28
PRIOR APPLICATION NUMBER: JP　　1998-62263
PRIOR FILING DATE: 1998-02-25
NUMBER OF SEQ ID NOS: 42
SOFTWARE: Patentin Ver. 2.1
SEQ ID NO 8
LENGTH: 594
TYPE: PRT
ORGANISM: Oryctolagus cuniculus
FEATURE:
NAME/KEY: misc_feature
LOCATION: 111
OTHER INFORMATION: Xaa is Met or Leu
US-10-139-262-8

Query Match　　14.4%;　Score 276.5;　DB 12;　Length 594;
Best Local Similarity　　24.9%;　Pred. No. 1e-19;
Matches　80;　Conservative　74;　Mismatches 120;　Indels 47;　Gaps 10;

```
Qy   72 RQLYVASAICLLFMIGEVVGGYLAHSLAVMTDAAHLLTDFASMLISLFSLWMSSRPATKT 131
         |!|: ||| |:| |  :|:|!|! |!|||:||!||  ||:  ||  ||  |!:
Db  246 RQIFYFLCLNLLFTFVELFYGVLTNSLGLISDGFHMLFDCSALVGLFAALMSRWKATRI 305

Qy  132 MNFGWQRAEILGALVSVLSIWVVTGVLVYLYLAVERLISGDYEIDGGTMLITSGCAVAVNII 191
         :!|: ||| |  :|: | |:|  | ||| |||
Db  306 FSYGYGRIEILSGFINGLFLIVIAFFVFMESVARLIDPP-ELDTHMLTPVSVGGLIVNLI 364

Qy  192 MGL---------------------TLHQS----------------------GHGHSHGTTNQOEENPSVRAAF 221
                                                      ||||||||| ::|:|
Db  365 -GICAFSHAHNHHGHSSQGSCHSSSDHSHSHHMHGHSDHGHGHSHGSPG-GGMNANMRGVF 422

Qy  222 IHVIGDFMQSMGVLV-AAYILYFKPEYKYVDPICTFVFSILVLGTTLTILRDVILVLMEG 280
         ||:|| : | |:|   ::||:|!||:||!:  ||:||| ::|!|  |!| :
Db  423 FHVLADTLGSIGVIVFTVFIEQF--GWFIADPLCSLFIAVLIFLSVVPLIKDACQVLLR 480

Qy  281 TPKGVDFTAVRDLLLSVEGVEALHSL-----HIWALTVAQPVLSVHIAIAQNTDAQAVL 334
              :!:| ||  |::   |:!| ||| ||| :|::
Db  481 LPPEYE---KELHIALEKIQKIEGLISYRDPHFWRHSASIVAGTIHIQVTSDVLEQRIV 536

Qy  335 KTASSRLQGKFHFHTVTIQIE 355
                |: :|||||!||:|
Db  537 QQVTGILKDA-GVNNLIIQVE 556
```

RESULT 25
US-10-139-262-10
Sequence 10, Application US/10139262
Patent No. US20020128459A1
GENERAL INFORMATION:
APPLICANT: NAKAMURA, YUSUKE
APPLICANT: TANAKA, TOSHIHIRO
APPLICANT: TSUKADA, SHUICHI
TITLE OF INVENTION: PHYSIOLOGICALLY ACTIVE PROTEIN ORIGINATING IN MAMMALS
FILE REFERENCE:
CURRENT APPLICATION NUMBER: US/10/139,262
CURRENT FILING DATE: 2002-05-07
PRIOR APPLICATION NUMBER: US/09/380,287A
PRIOR FILING DATE: 1999-11-16
PRIOR APPLICATION NUMBER: JP　　1997-62259
PRIOR FILING DATE: 1997-02-28
PRIOR APPLICATION NUMBER: JP　　1998-62263
PRIOR FILING DATE: 1998-02-25
NUMBER OF SEQ ID NOS: 42
SOFTWARE: Patentin Ver. 2.1
SEQ ID NO 10
LENGTH: 594
TYPE: PRT
ORGANISM: Homo sapiens
FEATURE:
NAME/KEY: misc_feature
LOCATION: 97
OTHER INFORMATION: Xaa is Phe or Leu
FEATURE:
NAME/KEY: misc_feature
LOCATION: 339
OTHER INFORMATION: Xaa is Lys or Arg
FEATURE:
NAME/KEY: misc_feature
LOCATION: 498
OTHER INFORMATION: Xaa is Lys or Glu
FEATURE:
NAME/KEY: misc_feature
LOCATION: 514, 528, 565
OTHER INFORMATION: Xaa is Phe or Ser
FEATURE:
NAME/KEY: misc_feature
LOCATION: 537
OTHER INFORMATION: Xaa is Gln or Arg
US-10-139-262-10

Query Match　　14.2%;　Score 272.5;　DB 12;　Length 594;
Best Local Similarity　　25.2%;　Pred. No. 2.6e-19;

Matches 80; Conservative 70; Mismatches 128; Indels 39; Gaps 9;

```
Qy    72 RQLYASAICLLFMIGEVVGGYLAHSLAVMTDAAHLLTDFASMLISLFSLWMSSRPATKT 131
         ::|||   |||   |  ||||  |   | ||| |::|||::|| ||||||||||
Db   246 RQIFYFLCLNLLFTEVLEFYGVLTNSIGLLISDGFHMLFDCSALVMGLFAALMSRWKATRI 305

Qy   132 MNFGWQRAEILGALVSVLSIWVVTGVLVLYLAVERLISGDYEIDGGTMLITSGCAVANII 191
         ||  | |   |||||| |  | |  | ||  | | || | |  ||  |||| ||||| |||
Db   306 FSYGYGRIEILSGFINGLFLIVIAFFVFMESVAXLIDPP-ELDTHMLTPVSVGGLIVNLI 364

Qy   192 -------------TLHQS---------------GHGHSHGTTNQQENPSVRAAF 221
                      |||||               ||||||| |   || |  ||
Db   365 -GICAFSHAHSHAGASQGSCHSSDHSHSHHMGHSDHGHGHSHGSAG--GGMNANMRGVF 422

Qy   222 IHVIGDFMQSMGVLVA-AYILYFKPEYKYVDPICTFVVFSILVLGTTLTLRDVILVLMEG 280
         |||  | |   |   | ||||| ||  |    |   |  :| |  |  | |  ||  |:|
Db   423 LHVIADTLGSIGVISTVFIEQF--GWFIADPLCSLFIAILIFLSVVPLIKDACQVILLR 480

Qy   281 TPK--GVDFTAVRDLLLSVEGVEALHSLHIWALTVAQPVLSVHIAIAQNTDAQAVLKTAS 338
            |     |    | |  |||| | | |   |  |  ||  |      | | || |
Db   481 LPPEYGKELHIALEKIQXIEGLISYRDPHFWRHXASIVAGTIHIQVTXDVLEQRIVXQVT 540

Qy   339 SRLQGKFHFHTVTIQIE 355
         ||     :  ||||  |
Db   541 GILKDA-GVNNLTIQVE 556
```

RESULT 26
US-10-012-542-233
; Sequence 233, Application US/10012542
; Publication No. US20030044851A1
; GENERAL INFORMATION:
; APPLICANT: Ruben et al.
; TITLE OF INVENTION: 94 Human Secreted Proteins
; FILE REFERENCE: PZ029P1
; CURRENT APPLICATION NUMBER: US/10/012,542
; CURRENT FILING DATE: 2001-12-12
; PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: 09/461,325
; PRIOR APPLICATION NUMBER: EARLIER FILING DATE: 1999-12-14
; PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: 60/089,507
; PRIOR FILING DATE: EARLIER FILING DATE: 1998-06-16
; PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: 60/089,508
; PRIOR FILING DATE: EARLIER FILING DATE: 1998-06-16
; PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: 60/089,509
; PRIOR FILING DATE: EARLIER FILING DATE: 1998-06-16
; PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: 60/089,510
; PRIOR FILING DATE: EARLIER FILING DATE: 1998-06-16
; PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: 60/090,112
; PRIOR FILING DATE: EARLIER FILING DATE: 1998-06-22
; PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: 60/090,113
; PRIOR FILING DATE: EARLIER FILING DATE: 1998-06-22
; SOFTWARE: Patentin Ver. 2.0
; SEQ ID NO 233
; LENGTH: 293
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: SITE
; LOCATION: (134)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
; US-10-012-542-233

Query Match   11.1%; Score 213.5; DB 9; Length 293;
Best Local Similarity   23.9%; Pred. No. 9.5e-14;
Matches   69; Conservative   48; Mismatches   65; Indels 107; Gaps   9;

```
Qy    79 AICLLFMIGEVVGGYLAHSLAVMTDAAHLLTDFASMLISLFSLWMSSRP-ATKTMNFGWQ 137
         ||:  || |||  |||  || ||  |||:|| |||::|:|::||||| | |   |||||
Db     3 ALTFMFMVLEVVVSRVTSSLAMLSDSFHMLSDVLALVVALVAERFARRTHATQKNTFGWI 62

Qy   138 RAEILGALVSVLSIWVVTGVLVLYLAVERLISGDYEIDGGTMLITSGCAVANIIMGLTL- 196
         ||:|  | | ||  ||| ||||        :  | |  |||||  |  |  |  ||| |
Db   138 RAEILGALVSVLSIWVVTGVLVLYLAVERLISGDYEIDGGTMLITSGCAVANIMGLTL- 196
```

```
Db  256 DALGSVIVVNALVFYF--SWKGCSEGDFCVNPCFPDPCKAFVEILIVL 302
```

RESULT 28
US-09-815-242-13500
Sequence 13500, Application US20020061569A1
Patent No. US20020061569A1
GENERAL INFORMATION:
- APPLICANT: Haselbeck, Robert
- APPLICANT: Ohlsen, Karl L.
- APPLICANT: Zyskind, Judith W.
- APPLICANT: Wall, Daniel
- APPLICANT: Trawick, John D.
- APPLICANT: Carr, Grant J.
- APPLICANT: Yamamoto, Robert T.
- APPLICANT: Xu, H. Howard
- TITLE OF INVENTION: Identification of Essential Genes in
     Prokaryotes
- FILE REFERENCE: ELITRA_011A
- CURRENT APPLICATION NUMBER: US/09/815,242
- CURRENT FILING DATE: 2001-03-21
- PRIOR APPLICATION NUMBER: 60/191,078
- PRIOR FILING DATE: 2000-03-21
- PRIOR APPLICATION NUMBER: 60/206,848
- PRIOR FILING DATE: 2000-05-23
- PRIOR APPLICATION NUMBER: 60/207,727
- PRIOR FILING DATE: 2000-05-26
- PRIOR APPLICATION NUMBER: 60/242,578
- PRIOR FILING DATE: 2000-10-23
- PRIOR APPLICATION NUMBER: 60/253,625
- PRIOR FILING DATE: 2000-11-27
- PRIOR APPLICATION NUMBER: 60/257,931
- PRIOR FILING DATE: 2000-12-22
- NUMBER OF SEQ ID NOS: 14110
- SOFTWARE: FastSEQ for Windows Version 4.0
- SEQ ID NO 13500
- LENGTH: 394
- TYPE: PRT
- ORGANISM: Streptococcus pneumoniae
US-09-815-242-13500

```
Query Match          6.9%;  Score 132;  DB 10;  Length 394;
Best Local Similarity  18.6%;  Pred. No. 2.6e-05;
Matches   56;  Conservative   69;  Mismatches 136;  Indels  40;  Gaps   8;

Qy  70 AQRQLYVASAICLLFMIGEVVGGYLAHSLAVMTDAAHLLTDFASMLISLFSLWMSSRPAT 129
Db  11 AERGAIISISTYLILSAAKLAAGHLLHSSSLVADGFNNVSDIIGNVALLIGIRMARQPAD 70

Qy 130 KTMNFGWQRAEILGALVSVLSIWVVTGVLVLAVERLISGDYE-ID--GGTMLITSGCAV 186
Db  71 RDHFEGHWKIEDLASLITSIIMFYVGFDVLRDTIQKILSREESVIDFLGATLGIISAAIM 130

Qy 187 AVNIIMGLTLHQSGHGHSHGTTNQQEENPSVRAAFIHVIGDFMQSMGVLVAAYILYFKPE 246
Db 131 FVVYLYNTRL----------SKKSNSKALKAAAKDNLSDAVTSLGTAIA-ILASSFN 176

Qy 247 YKYVDPICTFVFSILVLGTTLTILRDVILVLMEGTPKGVDFTAVRD---LLLSVEGVEAL 303
Db 177 YPIVDKLVAIITFFLKTAVDIFIESSFSLSD----GFDDRLLEDYQKAIMEIPKISKV 232

Qy 304 HS--------LHIWALTVAQPVLSVHIAIAQNTDAQAVLKTASSRLQGKFHFHTVTIQI 354
Db 233 KSQRGRTYGSNIYLDITLEMNPDLSVP-----ESHEIADQVESMLENRFGVFDTDVHI 285

Qy 355 E 355
Db 286 E 286
```

RESULT 29

US-09-738-626-6557
Sequence 6557, Application US/0973862A1
Publication No. US20020197605A1
GENERAL INFORMATION:
- APPLICANT: NAKAGAWA, SATOSHI
- APPLICANT: MIZOGUCHI, HIROSHI
- APPLICANT: ANDO, SEIKO
- APPLICANT: HAYASHI, MIKIRO
- APPLICANT: OCHIAI, KEIKO
- APPLICANT: YOKOI, HARUHIKO
- APPLICANT: TATEISHI, NAOKO
- APPLICANT: SENOH, AKIHIRO
- APPLICANT: IKEDA, MASATO
- APPLICANT: OZAKI, AKIO
- TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
- FILE REFERENCE: 249-125
- CURRENT APPLICATION NUMBER: US/09/738,626
- CURRENT FILING DATE: 2000-12-18
- PRIOR APPLICATION NUMBER: JP 99/377484
- PRIOR FILING DATE: 1999-12-16
- PRIOR APPLICATION NUMBER: JP 00/159162
- PRIOR FILING DATE: 2000-04-07
- PRIOR APPLICATION NUMBER: JP 00/280988
- PRIOR FILING DATE: 2000-08-03
- NUMBER OF SEQ ID NOS: 7059
- SOFTWARE: Patentin ver. 3.0
- SEQ ID NO 6557
- LENGTH: 286
- TYPE: PRT
- ORGANISM: Corynebacterium glutamicum
US-09-738-626-6557

```
Query Match          6.2%;  Score 120;  DB 9;  Length 286;
Best Local Similarity  22.0%;  Pred. No. 0.00028;
Matches   62;  Conservative   54;  Mismatches 126;  Indels  40;  Gaps   8;

Qy  72 RQLYVASAICLLFMIGEVVGGYLAHSLAVMTDAAHLLTDFASMLISLFSLW----MSSRP 127
Db   9 RALKFSLVLMLLTTVLQAIIVSFSGGSVALLADTVHNLSD----ALTAIPLWIAFILSRRA 64

Qy 128 ATKTMNFGWQRAEILGALVSVLSIWVVTGVLVLAVERLISG------DYEIDGGTMLIT 181
Db  65 ATQKYTYGFNRAEDLAGLFIVAMIALSAIVAAWQAIDRMINPRPMENIEWVIAAGVIGFL 124

Qy 182 SGCAVAVNIIMGLTLHQSGHGHSHGTTNQQEENPSVRAAFIHVIGDFMQSMGVLVAAYIL 241
Db 125 GNEAVAMTRI-------------RVGKRIGSAALVADGVHARTDGFTSLAVVAGGVGV 169

Qy 242 YFKPEYKYVDPICTFVFSILVLGTTLTILRDVILVLMEGTPKGVDFTAVRDL--LLSVEG 299
Db 170 FL--GFPLADPIGLIISAMIATLLVGTIRSVGRRLMDGIEPELVEKAPHAIWHVKEIS 227

Qy 300 VEAL------HSLHIWALTVAQPVLSVHIAIAQNTDAQAVLK 335
Db 228 IDRLRLRWVGHRLHGDA-TVSTSTSSLSEATAIALEAELSVK 268
```

RESULT 30
US-09-823-114-21
Sequence 21, Application US/09823114
Patent No. US20020061554A1
GENERAL INFORMATION:
- APPLICANT: EVANS, CHRISTOPHER J.
- APPLICANT: KEITH, DUANE E.
- TITLE OF INVENTION: OPIOID RECEPTOR GENES
- NUMBER OF SEQUENCES: 25
- CORRESPONDENCE ADDRESS:
- ADDRESSEE: MORRISON & FOERSTER
- STREET: 2000 PENNSYLVANIA AVENUE, NW, Suite 5500
- CITY: WASHINGTON
- STATE: DC
- COUNTRY: USA
- ZIP: 20006-1888

COMPUTER READABLE FORM:
  MEDIUM TYPE: Floppy disk
  COMPUTER: IBM PC compatible
  OPERATING SYSTEM: PC-DOS/MS-DOS
  SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
  APPLICATION NUMBER: US/09/823,114
  FILING DATE: 29-Mar-2001
  CLASSIFICATION: <Unknown>
PRIOR APPLICATION DATA:
  APPLICATION NUMBER: 09/148,351
  FILING DATE: <Unknown>
ATTORNEY/AGENT INFORMATION:
  NAME: MURASHIGE, KATE H.
  REGISTRATION NUMBER: 29,959
  REFERENCE/DOCKET NUMBER: 22000-20526.22
TELECOMMUNICATION INFORMATION:
  TELEPHONE: (202) 887-1500
  TELEFAX: (202) 887-0763
  TELEX: 90-4030 MRSNFOERSWSH
INFORMATION FOR SEQ ID NO: 21:
SEQUENCE CHARACTERISTICS:
  LENGTH: 370 amino acids
  TYPE: amino acid
  STRANDEDNESS: single
  TOPOLOGY: linear
FEATURE:
  NAME/KEY: Modified-site
  LOCATION: group(18, 33)
  OTHER INFORMATION: /note= "extracellular Asn residues
    that are consensus sites for N-linked glycosylation"
SEQUENCE DESCRIPTION: SEQ ID NO: 21:
US-09-823-114-21

Query Match   5.2%;   Score 100.5;   DB 10;   Length 370;
Best Local Similarity 21.2%;   Pred. No. 0.038;
Matches 83;   Conservative 59;   Mismatches 137;   Indels 113;   Gaps 20;

Qy   32 PLPRPGDLDLQAIELAAQSNHHCHA--QKGPDSHCDPKKGKAQRQLYVA------SAICLLF 84
             ::|| |: ||| :: | |: |:|| || ||:| | | | ::|
Db   3 PAPSAGAELQPFLFANASDAYPSACPSAGANASGPPARSASSLALAIATTALYSAVCAVG 62

Qy   85 MIGEVVGGYLAHSLAVMTDAAHLLTDFASMLISLFSLWMSSRPATKTMNF--------GW 136
             |: | | | |: :| ||| || |||::| |:| | | | ||
Db   63 LIGNVL-------------VMFGIVRYTKMKTATNIYIFNLALADALATSTLPFQSAKYLMETW 113

Qy 137 QRAEIL-GALVSV--------LSIWVVTGVLV--YLAVERLISG-DYEIDGGTMLI------ 180
             ::|| |:| | |: || ||:| :| | |: : |:| |: ::
Db 114 PFGELLCKAVLSIDYNMFTSIFTLIMMSVDRYIAVCHPVKALDFRTPAKAKLINICIWV 173

Qy 181 -TSGCAAVNIIMGLLTLHQSGHGHSHGTTNQQEENPSVRAAFIHVIGDFMQSMGVLVAAY 239
              ::| ||: : |:| ||||| | ::||||| || || :| ||::|| ||||:|
Db 174 LASGVGVPI-MVMAVTRPRDG-------------------------------AVVCM 198

Qy 240 ILYFKPEYKY--VDPICTFVFSILVLGTTLTILRDVILV-------LMEGT-PKGVDFTAV 290
             ||:| | |||:|| |:|||||||| ||:|:| :| : |||| |
Db 199 LQFPSPSWYWDTVTKICVFLFAFVVPILIITVCYGLMLLRLRSVRLLSGSKEKDRSLRRI 258

Qy 291 RDLLSVEG--VEALHSLHI------WALTVAQ---PVLSVHIAIAQNTDAQAVLKTASSR 340
             | :| ||:| | || |:: ||:|::|| |::| : |:|:::
Db 259 TRMVLVVGAFVVCWAPIHFVIVWTLVDIRRDPLVVAALHLCIA-----LGTANSS 310

Qy 341 LQGKFHFHTVTQIEDYSEDMKDC--OACQGP 370
             | :| | ||:|:| :: || ||| |
Db 311 LN-------PVLYAFLDENFKRCFRQLCRKP 334

APPLICANT: MIZOGUCHI, HIROSHI
APPLICANT: ANDO, SEIKO
APPLICANT: HAYASHI, MIKIRO
APPLICANT: OCHIAI, KEIKO
APPLICANT: YOKOI, HARUHIKO
APPLICANT: TATEISHI, NAOKO
APPLICANT: SENOH, AKIHIRO
APPLICANT: IKEDA, MASATO
APPLICANT: OZAKI, AKIO
TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
FILE REFERENCE: 249-125
CURRENT APPLICATION NUMBER: US/09/738,626
CURRENT FILING DATE: 2000-12-18
PRIOR APPLICATION NUMBER: JP 99/377484
PRIOR FILING DATE: 1999-12-16
PRIOR APPLICATION NUMBER: JP 00/159162
PRIOR FILING DATE: 2000-04-07
PRIOR APPLICATION NUMBER: JP 00/280988
PRIOR FILING DATE: 2000-08-03
NUMBER OF SEQ ID NOS: 7059
SEQ ID NO 3686
SOFTWARE: Patentin ver. 3.0
SEQ ID NO 3686
LENGTH: 458
TYPE: PRT
ORGANISM: Corynebacterium glutamicum
US-09-738-626-3686

Query Match   5.2%;   Score 99.5;   DB 9;   Length 458;
Best Local Similarity 20.5%;   Pred. No. 0.064;
Matches 73;   Conservative 52;   Mismatches 104;   Indels 127;   Gaps 18;

Qy   5 EKQHLLDA----------RPAIRSYTGSLWQEGAGW--------------IP---------LPR 35
             :: ||| : |: ||:
Db 133 DPKDLLDAIAFLTWPALVAPVIAPLLGGLLADTIGWRNIFLLNVPLGIIAIIAGLFIPLPK 192

Qy   36 -------PG-----LDLQAIELAAQSNHHCHAQKGPDSHCDPKKGKAQRQLYVASA 79
                   |: |||| | : | |: | ||:||
Db 193 NTAVNVKRFDLPGFLGAMLVMVALTVAAEL-------ISRGSSAE----------LTIA 234

Qy   80 ICLLFMIGEVVGGYLAHSLAVMTDAAHLLTDFASMLI------SLFSLWMSSRPAT 129
             ||||:| |:| ||| | || ||||||||||||:: :   :|| | ||
Db 235 ACLV-LSAAVVCGFVVRNLRV---PGRLFDLSIMRIPGFRVGNSSGSSIYRLVITAAPFM 289

Qy 130 KTM----NFGWQRAEILGALV-------------SVLSIWVVTGVLYLAVERLISGD 170
             |: | :||| ||: |: ||| || |:| |::|
Db 290 FTLLFQVAFGWS-ATLAGAMVVALFAGNVAIKPFTTPIIKRWNFKPVLVFSNA------ 341

Qy 171 YEIDGGTMLITSGCAVAVN---IIMGLLTLHQSGHGHSHG------------TTNQQEENPS 216
             :::| |::| || | :| | | |||||| | :   :||:| ||
Db 342 ----AGALVLATFLFVRADTPLVLIVLLLFVSGALRSLGFSAYNTLQFVDISPEQTSNAN 397

Qy 217 VRAAFIHVIGDFMQSMGVLVAAXILYFKPEYKYVDPICTFVFSILVLGTTLTILRD 272
             |:| |:|:|| : |: || : || |:|||:| |:|:|||:||:||||| ||||
Db 398 VLSATLHOLG--MSLGIAVAVIAMSLAPTANWAFPLAAALFLIPLIG-ALSLPRD 449

RESULT 31
US-09-738-626-3686
Sequence 3686, Application US/09738626
Publication No. US20020197605A1
GENERAL INFORMATION:
APPLICANT: NAKAGAWA, SATOSHI

RESULT 32
US-09-738-626-5835
Sequence 5835, Application US/09738626
Publication No. US20020197605A1
GENERAL INFORMATION:
APPLICANT: NAKAGAWA, SATOSHI
APPLICANT: MIZOGUCHI, HIROSHI
APPLICANT: ANDO, SEIKO
APPLICANT: HAYASHI, MIKIRO
APPLICANT: OCHIAI, KEIKO
APPLICANT: YOKOI, HARUHIKO
APPLICANT: TATEISHI, NAOKO
APPLICANT: SENOH, AKIHIRO
APPLICANT: IKEDA, MASATO
APPLICANT: OZAKI, AKIO
TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
FILE REFERENCE: 249-125

```
CURRENT APPLICATION NUMBER: US/09/738,626
CURRENT FILING DATE: 2000-12-18
PRIOR APPLICATION NUMBER: JP 99/377484
PRIOR FILING DATE: 1999-12-16
PRIOR APPLICATION NUMBER: JP 00/159162
PRIOR FILING DATE: 2000-04-07
PRIOR APPLICATION NUMBER: JP 00/280988
PRIOR FILING DATE: 2000-08-03
NUMBER OF SEQ ID NOS: 7059
SOFTWARE: Patentin ver. 3.0
SEQ ID NO 5835
LENGTH: 306
TYPE: PRT
ORGANISM: Corynebacterium glutamicum
US-09-738-626-5835

Query Match
Best Local Similarity  20.0%;  Score 94.5;  DB 9;  Length 306;
Matches  62;  Conservative  53;  Mismatches 114;  Indels  81;  Gaps  12;

Qy   76 VASAICLLFMIGEVVGGYLAHSLAVMTDAAHLLTDFASMLISLFSLWMSSRPATKTMNFG 135
Db   13 IAWVLFSIVLVGVVAQGSQAMRSAWIEDMLSLVPPIAFLLASR------ISKAVATRKHPYG 68

Qy  136 WQRAEILGALVSVLSIWVVTGVLVLYLAVERLISGD----------YEIDGGTMLITSGC 184
Db   69 KHRSIAIGHQAAALALLIMGSLLIYEAVSALIKGERPPIGILAILFGHDVWSGWLMI--GV 126

Qy  185 AVAVNIIMGLI---------LHQSGHGHSHGTTNQQEENPSVRAAFIHVIGDFMQS-MG 233
Db  127 MISVSIPMVIGVRVKINLAKDLHDK----------LLIYADADMAKADWG 165

Qy  234 VLVAAY--ILYFKPEYKYVDPICTFVFSILVLGTTLTILRDVILVLMEGTPKGVDFTAVR 291
Db  166 TAVASIVGVLGIGLGFWWADAVAAALVISASILRDGVINMKAAISDLGDGRAMTYDNSAPH 225

Qy  292 DLLLSVEGVEALHSLHIWALTVAQPVLSVHIAIAQNTDAQAVLKTASSRL--QGK-FHFH 348
Db  226 PLNDDVE--KAALEMH-W------------------VKHARARIRDQGRCFHTE 258

Qy  349 TVTIQIEDYS 358
Db  259 LFVEPVEGYT 268


RESULT 33
US-10-112-599A-4
Sequence 4, Application US/10112599A
Publication No. US20030005476A1
GENERAL INFORMATION:
APPLICANT: Allen, Keith D.
TITLE OF INVENTION: DELTA OPIOID RECEPTOR DISRUPTIONS,
TITLE OF INVENTION: COMPOSITIONS AND METHODS RELATING THERETO
FILE REFERENCE: R-678
CURRENT APPLICATION NUMBER: US/10/112,599A
CURRENT FILING DATE: 2002-09-05
PRIOR APPLICATION NUMBER: US 60/280,513
PRIOR FILING DATE: 2001-03-29
NUMBER OF SEQ ID NOS: 6
SOFTWARE: FastSEQ for Windows Version 4.0
SEQ ID NO 4
LENGTH: 372
TYPE: PRT
ORGANISM: Homo sapiens
US-10-112-599A-4

Query Match
Best Local Similarity  21.3%;  Score 93.5;  DB 9;  Length 372;
Matches  84;  Conservative  57;  Mismatches 138;  Indels 115;  Gaps  21;

Qy   32 PLPRPGLDLQAIELAAQSNHHCHA--QKGPDSHCDPKKGKAQR-QLYVA------SAICLL 83
Db    3 PAPSAGAELQPPLFANASDAYPSAFPSAGANASGPPGPGSASSLALAIAITALYSAVCAV 62
```

```
RESULT 34
US-09-738-626-5934
Sequence 5934, Application US/09738626
Publication No. US20020197605A1
GENERAL INFORMATION:
APPLICANT: NAKAGAWA, SATOSHI
APPLICANT: MIZOGUCHI, HIROSHI
APPLICANT: ANDO, SEIKO
APPLICANT: HAYASHI, MIKIRO
APPLICANT: OCHIAI, KEIKO
APPLICANT: YOKOI, HARUHIKO
APPLICANT: TATEISHI, NAOKO
APPLICANT: SENOH, AKIHIRO
APPLICANT: IKEDA, MASATO
APPLICANT: OZAKI, AKIO
TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
FILE REFERENCE: 249-1125
CURRENT APPLICATION NUMBER: US/09/738,626
CURRENT FILING DATE: 2000-12-18
PRIOR APPLICATION NUMBER: JP 99/377484
PRIOR FILING DATE: 1999-12-16
PRIOR APPLICATION NUMBER: JP 00/159162
PRIOR FILING DATE: 2000-04-07
PRIOR APPLICATION NUMBER: JP 00/280988
PRIOR FILING DATE: 2000-08-03
NUMBER OF SEQ ID NOS: 7059
SOFTWARE: Patentin ver. 3.0
SEQ ID NO 5934
LENGTH: 441
TYPE: PRT
ORGANISM: Corynebacterium glutamicum
US-09-738-626-5934

Query Match
Best Local Similarity  23.9%;  Score 93;  DB 9;  Length 441;
Matches  65;  Conservative  47;  Mismatches  90;  Indels  70;  Gaps  16;

Qy   84 FMIGEVVGGYLAHSLAVMTDAAHLLTDFASMLISLFSLWMSSRPATKTMNF--------G 135
Db   63 GLLGNVL------------VMFGIVRYTKMKTATNIYIFNLALADALATSTLPFQSAKYLMET 113

Qy  136 WQRAEIL-GALVSV------LSIWVVTGVLV--YLAVERLISG-DYEIDGGTMLIT----- 180
Db  114 WPFGELLCKAVLSIDYYNMFTSIFTLTMMSVDRYIAVCHPVKALDFRPAKAKLINICIW 173

Qy  181 --TSGCAVAVNIIMGLTLHQSGHGHSHGTTNQQEENPSVRAAFIHVIGDFMQSMGVLVAA 238
Db  174 VLASGVGVPI-MVMAVTRPRDG-------------------------AVVC 198

Qy  239 YILYFKPEYKY--VDPICTFVFSILVLGTTLTILRDVILV--------LMEGT-PKGVDFTA 289
Db  199 MLQFPSPSMYWDIVTKICVFLFAFVVPILIITVCYGLMLLRLRSVRLLSGSKEKDRSLRR 258

Qy  290 VRDLLLSVEG--VEALHSLHI-----WALTVAQ-----PVLSVHIAIAQNTDAQAVLKTAS 338
Db  259 ITRMVLVVVGAFVVCWAPIHIFVIVWTLVDIDRRDPLVVAALHCIA------LGYAN 310

Qy  339 SRLQGKFHFHTVTIQIEDYSEDMKDC--QACQGP 370
Db  311 SSLN--------PVLIYAFLDENFKRCFRQLCRKP 336


Query Match
Best Local Similarity  23.9%;  Score 93;  DB 9;  Length 441;
Matches  65;  Conservative  47;  Mismatches  90;  Indels  70;  Gaps  16;

Qy   32 PLPRPGLDLQAIELAAQSNHHCHAQKGPDSHCDPKKGKAQRQLYVASAIC------LLF 84
Db  205 PQPKP-----VESAQEA------------DKPSLRASLKLIVRNPTTWQGFFIHYVL 244

Qy   85 MIGEVVGGYLAHSLAVMTDAAHLLTDFASMLISLFSLWM-SSRPATKTMN--FGWQRAEI 141
Db  245 MVWQTVFS-MMWGVPLMTLGMGLSATTAGLVLSINTLCMVVSAPIIGIISARLGYRRDV 303

Qy  142 LGALVSVLS-IWVVTGVLVLYLAVERLISGDYEIDGGTMLITSGCAVAVNIIMGLTLHQSG 200
```

```
Db   304  AIALSFVQSAVW------LVFLA------SDAPRGLMAI-----ILVNIVMGLTTAASG 345
Qy   201  HGHSHGTTNQQEENPSVRAAFIHV--IGDFMQSMGVLVAAYILYFKPE------YKYVD 251
Db   346  YGFD---TIRERLDRKILAAGTGLANMGGFLSSM---VAAQVMGFLLDHSAHGSTYTWVD 399
Qy   252  PICTFVFSILVLGTTLTI-LRDVILVLMEGTP 282
Db   400  ---FRFGFLAILVTWAVGVTGFVVARLKGGP 427
```

RESULT 35
US-09-510-332-77
; Sequence 77, Application US/09510332
; Publication No. US20030022278A1
; GENERAL INFORMATION:
;   APPLICANT: Zuker, Charles S.
;   APPLICANT: Adler, Jon Elliot
;   APPLICANT: Ryba, Nick
;   APPLICANT: Mueller, Ken
;   APPLICANT: Hoon, Mark
;   APPLICANT: The Regents of the University of California
;   TITLE OF INVENTION: T2R, a No. US20030022278A1el Family of Taste Receptors
;   FILE REFERENCE: 02307E-098010US
;   CURRENT APPLICATION NUMBER: US/09/510,332
;   CURRENT FILING DATE: 2000-02-22
;   PRIOR APPLICATION NUMBER: US 09/393,634
;   PRIOR FILING DATE: 1999-09-10
;   NUMBER OF SEQ ID NOS: 172
;   SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 77
;   LENGTH: 335
;   TYPE: PRT
;   ORGANISM: Rattus sp.
;   FEATURE:
;   OTHER INFORMATION: rat T2R01 (rGR01)
US-09-510-332-77

Query Match   4.7%; Score 91; DB 9; Length 335;
Best Local Similarity 19.9%; Pred. No. 0.3;
Matches 59; Conservative 44; Mismatches 92; Indels 102; Gaps 16;

```
Qy    80  ICLLFMIGEVVGGYLAHSLAVMTDAAHLLTD------FASMLISL-FSLWMSSRPATKTM 132
          :|:||                                 ||:||: |:
Db    60  LCILFA--------QLCLFSLVRHTLFEDNITFVFIINELSLWFATWLGVFYCAKIA 108

Qy   133  N-----FGWQRAEILGALVSVLSIWVVTGVLVLYLAVERLISGDYEIDGGTML------- 179
          |:|:   ||
Db   109  TIPHPLFWLKMRI----SRLVPWLILGSVLVIVITTFI---HSRETSAILKPIFISLF 160

Qy   180  -----ITSGCAVANI-IMGLTL----------------------------HQSGHG 202
          ||:|: ::|:||||
Db   161  PKNATQVGTGHATLLSVLVLGLTLPLFIFTVAVLLLIYSLMWNYSRQMRTMVGTREYSGHA 220

Qy   203  HSHGTTNQQEENPSVRAAFIHVIGDFMQSMGVLVAAYILYFKPEYKYVDPICTFVFSILV 262
          |||  |:|
Db   221  HISAML-------SILSFLIYLSHYM--VAVLISTQVLYLGSR------TEVFCLLV 263

Qy   263  LGTTLTILLRDVILVLMEGTPKGVDFTAVRDLLLSVEGEVEALHSLHIWALTVAQPVLS 319
          :|: :|:  |:
Db   264  IGMYPSI-HSIVLIL--GNPK----LKRNAKMFIVHCKCCHCTRAW-VISRSPRLS 311
```

;   APPLICANT: The Regents of the University of California
;   APPLICANT: The Government of the United States of America
;   APPLICANT: as represented by the Secretary of the
;   APPLICANT: Department of Health and Human Services
;   TITLE OF INVENTION: SF, a No. US20020051997A1el Family of Taste Receptors
;   FILE REFERENCE: 023078-098000US
;   CURRENT APPLICATION NUMBER: US/09/393,634
;   CURRENT FILING DATE: 1999-09-10
;   NUMBER OF SEQ ID NOS: 92
;   SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 1
;   LENGTH: 335
;   TYPE: PRT
;   ORGANISM: Rattus sp.
;   FEATURE:
;   OTHER INFORMATION: rat GR01, sF01
US-09-393-634-1

Query Match   4.7%; Score 91; DB 10; Length 335;
Best Local Similarity 19.9%; Pred. No. 0.3;
Matches 59; Conservative 44; Mismatches 92; Indels 102; Gaps 16;

```
Qy    80  ICLLFMIGEVVGGYLAHSLAVMTDAAHLLTD------FASMLISL-FSLWMSSRPATKTM 132
          :|:||                                 ||:||: |:
Db    60  LCILFA--------QLCLFSLVRHTLFEDNITFVFIINELSLWFATWLGVFYCAKIA 108

Qy   133  N-----FGWQRAEILGALVSVLSIWVVTGVLVLYLAVERLISGDYEIDGGTML------- 179
          |:|:   ||
Db   109  TIPHPLFWLKMRI----SRLVPWLILGSVLVIVITTFI---HSRETSAILKPIFISLF 160

Qy   180  -----ITSGCAVANI-IMGLTL----------------------------HQSGHG 202
          ||:|: ::|:||||
Db   161  PKNATQVGTGHATLLSVLVLGLTLPLFIFTVAVLLLIYSLMWNYSRQMRTMVGTREYSGHA 220

Qy   203  HSHGTTNQQEENPSVRAAFIHVIGDFMQSMGVLVAAYILYFKPEYKYVDPICTFVFSILV 262
          |||  |:|
Db   221  HISAML-------SILSFLIYLSHYM--VAVLISTQVLYLGSR------TEVFCLLV 263

Qy   263  LGTTLTILLRDVILVLMEGTPKGVDFTAVRDLLLSVEGEVEALHSLHIWALTVAQPVLS 319
          :|: :|:  |:
Db   264  IGMYPSI-HSIVLIL--GNPK----LKRNAKMFIVHCKCCHCTRAW-VISRSPRLS 311
```

RESULT 37
US-09-966-871-80
; Sequence 80, Application US/09966871
; Patent No. US20020127539A1
; GENERAL INFORMATION:
;   APPLICANT: Koplin, Alan S.
;   TITLE OF INVENTION: Assays for Identifying Receptors Having
;   TITLE OF INVENTION: Alterations in Signaling
;   FILE REFERENCE: 00398/512002
;   CURRENT APPLICATION NUMBER: US/09/966,871
;   CURRENT FILING DATE: 2001-09-28
;   PRIOR APPLICATION NUMBER: US 60/236,302
;   PRIOR FILING DATE: 2000-09-28
;   PRIOR APPLICATION NUMBER: US 60/288,644
;   PRIOR FILING DATE: 2001-05-03
;   NUMBER OF SEQ ID NOS: 87
;   SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 80
;   LENGTH: 372
;   TYPE: PRT
;   ORGANISM: Homo sapiens
US-09-966-871-80

Query Match   4.7%; Score 89.5; DB 10; Length 372;
Best Local Similarity 20.3%; Pred. No. 0.5;
Matches 80; Conservative 60; Mismatches 139; Indels 115; Gaps 20;

RESULT 36
US-09-393-634-1
; Sequence 1, Application US/09393634
; Patent No. US20020051997A1
; GENERAL INFORMATION:
;   APPLICANT: Zuker, Charles S.
;   APPLICANT: Adler, Jon Elliot
;   APPLICANT: Ryba, Nick
;   APPLICANT: Mueller, Ken
;   APPLICANT: Hoon, Mark

```
Qy    32  PLPRPGLDLQAIELAAQSNHHCHAQKGPDSHCD-PKKGKAQRQLYVA----------SAICLL 83
          |:|| :                                                    ||:|:
Db     3  PAPSAGAELQPPLFANASDAYPSACPSAGANASGPPGARSASSLALAITALYSAVCAV 52
```

```
Qy    84  FMIGEVVGGYLAHSLAVMTDAAHLLTDFASMLISLFSLWMSSRPATKTMNF------G  135
          ::|: |:|
Db    63  GLLGNVL-----------VMFGIVRYTKMKTATNIYIFNLALADALATSTLPFQSAKYLMET  113

Qy   136  WQRAEIL-GALVSV-----LSIWVTGVLV--YLAVERLISG-DYEIDGGTMLI----  180
          || |:||:
Db   114  WPFGELLCKAVISIDYYNMFTSIFTLTMMSVDRYIAVCHPVKALDFRTPAKAKIINICIW  173

Qy   181  --TSGCAVANIIMGILTLHQSGHGHSHGTTNQQEENPSVRAAFIHVIGDFMQSMGVLVAA  238
          ::||| |:|::
Db   174  VLASGVGVPI-MVMAVTRPRDG-----------------------------AVVC  198

Qy   239  YILYFKPEYKY--VDPICTFVPSILVLGTTLTLILRDVILV-----LMEGT-PKGVDFTA  289
          :||
Db   199  MLQFPSPSWYWDTVTKICVFLFAFVVPILIITVCYGLMLLRSVRLLSGSKEKDRSLRR  258

Qy   290  VRDLLLSVEG--VEALHSLHI----WALTVAQ----PVLSVHIAIAQNTDAQAVLKTAS  338
          |: ||: |
Db   259  ITRMVLVVVGAFVVCWAPIHIFVIVWTLVDIDRRDPLVVAALHLCIA------LGYAN  310

Qy   339  SRLQGKFHFHTVTIQIEDYSEDMKDC--QACQGP  370
          :|: | |
Db   311  SSLN-------PVLYAFLDENFKRCFRQLCRKP  336
```

RESULT 38
US-10-039-645-80
Sequence 80, Application US/10039645
Patent No. US2002014717OA1
GENERAL INFORMATION:
;   APPLICANT: Kopin, Alan S.
;   APPLICANT: Beinborn, Martin
;   TITLE OF INVENTION: Constitutively Active, Hypersensitive,
;   TITLE OF INVENTION: and No. US20020147170A1functional Receptors as No. US20020147170
;   FILE REFERENCE: 0398/510002
;   CURRENT APPLICATION NUMBER: US/10/039,645
;   CURRENT FILING DATE: 2001-10-25
;   PRIOR APPLICATION NUMBER: US 60/243,550
;   PRIOR FILING DATE: 2000-10-26
;   NUMBER OF SEQ ID NOS: 87
;   SOFTWARE: FastSEQ for Windows Version 4.0
;   SEQ ID NO 80
;   LENGTH: 372
;   TYPE: PRT
;   ORGANISM: Homo sapiens
US-10-039-645-80

Query Match   4.7%;   Score 89.5;   DB 12; Length 372;
Best Local Similarity 20.3%;   Pred. No. 0.5;
Matches 80; Conservative 60; Mismatches 139; Indels 115; Gaps 20;

```
Db   259  ITRMVLVVVGAFVVCWAPIHIFVIVWTLVDIDRRDPLVVAALHLCIA------LGYAN  310
Qy   339  SRLQGKFHFHTVTIQIEDYSEDMKDC--QACQGP  370
          :|: | |
Db   311  SSLN-------PVLYAFLDENFKRCFRQLCRKP  336
```

RESULT 39
US-09-815-242-10236
Sequence 10236, Application US/09815242
Patent No. US20020615669A1
GENERAL INFORMATION:
;   APPLICANT: Haselbeck, Robert
;   APPLICANT: Ohlsen, Kari L.
;   APPLICANT: Zyskind, Judith W.
;   APPLICANT: Wall, Daniel
;   APPLICANT: Trawick, John D.
;   APPLICANT: Carr, Grant J.
;   APPLICANT: Yamamoto, Robert T.
;   APPLICANT: Xu, H. Howard
;   TITLE OF INVENTION: Identification of Essential Genes in
;   TITLE OF INVENTION: Prokaryotes
;   FILE REFERENCE: ELITRA.011A
;   CURRENT APPLICATION NUMBER: US/09/815,242
;   CURRENT FILING DATE: 2001-03-21
;   PRIOR APPLICATION NUMBER: 60/191,078
;   PRIOR FILING DATE: 2000-03-21
;   PRIOR APPLICATION NUMBER: 60/206,848
;   PRIOR FILING DATE: 2000-05-23
;   PRIOR APPLICATION NUMBER: 60/207,727
;   PRIOR FILING DATE: 2000-05-26
;   PRIOR APPLICATION NUMBER: 60/242,578
;   PRIOR FILING DATE: 2000-10-23
;   PRIOR APPLICATION NUMBER: 60/253,625
;   PRIOR FILING DATE: 2000-11-27
;   PRIOR APPLICATION NUMBER: 60/257,931
;   PRIOR FILING DATE: 2000-12-22
;   PRIOR APPLICATION NUMBER: 60/269,308
;   PRIOR FILING DATE: 2001-02-16
;   NUMBER OF SEQ ID NOS: 14110
;   SOFTWARE: FastSEQ for Windows Version 4.0
;   SEQ ID NO 10236
;   LENGTH: 412
;   TYPE: PRT
;   ORGANISM: Escherichia coli
US-09-815-242-10236

Query Match   4.6%;   Score 89;   DB 10; Length 412;
Best Local Similarity 19.9%;   Pred. No. 0.64;
Matches 80; Conservative 53; Mismatches 116; Indels 154; Gaps 18;

```
Qy    32  PLPRPGLDLQAIELAAQSNHHCHAQKGPDSHCD-PKKGKAQRQLYVA--------SAICLL  83
Db     3  PAPSAGAELQPLFANASDAYPSACPSAGANASGPPGARSASSIALAIAITRALYSAVCAV  62

Qy    84  FMIGEVVGGYLAHSLAVMTDAAHLLTDFASMLISLFSLWMSSRPATKTMNF------G  135
Db    63  GLLGNVL-----------VMFGIVRYTKMKTATNIYIFNLALADALATSTLPFQSAKYLMET  113

Qy   136  WQRAEIL-GALVSV-----LSIWVTGVLVL--YLAVERLISG-----DYEID------G  175
Db   124  LGVSLLQGAVLTGIATFLFLILMLQRRGQKPLEKVIGGLLLFVAAAYIVELIFSQPNLAQLG  183

Qy   176  GTMLI-----TSGCAVANIIMGILT------LHQSGHGHSHGTTNQQEENPSV-RAAFIH  223
Db   184  KGMVIPSLPTSEAVFLAAGVLGATIMPHVIYLHSSLTQHLHGGSRQQRYSATKWDVAIM  243

Qy   224  VIGDFMQSMGVLVAAYILYFK------PEYKYVDPICTFV------FSIL  261
Db   244  TIAGFVNLAMMATAAAAFHFSGHTGVADDLDEAYLTLQPLLSHAAATVFGLSLVAAGLSST  303

Qy   262  VLGT------------------------TLTILRDVILVLMEGTPKGVDFTAVRDLLLSVEGV  300
```

```
RESULT 40
US-09-738-626-4772
; Sequence 4772, Application US/09738626
; Publication No. US20020197605A1
; GENERAL INFORMATION:
; APPLICANT: NAKAGAWA, SATOSHI
; APPLICANT: MIZOGUCHI, HIROSHI
; APPLICANT: ANDO, SEIKO
; APPLICANT: HAYASHI, MIKIRO
; APPLICANT: OCHIAI, KEIKO
; APPLICANT: YOKOI, HARUHIKO
; APPLICANT: TATEISHI, NAOKO
; APPLICANT: SENOH, AKIHIRO
; APPLICANT: IKEDA, MASATO
; APPLICANT: OZAKI, AKIO
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
; FILE REFERENCE: 249-125
; CURRENT APPLICATION NUMBER: US/09/738,626
; CURRENT FILING DATE:   2000-12-18
; PRIOR APPLICATION NUMBER: JP 99/377484
; PRIOR FILING DATE: 1999-12-16
; PRIOR APPLICATION NUMBER: JP 00/159162
; PRIOR FILING DATE: 2000-04-07
; PRIOR APPLICATION NUMBER: JP 00/280988
; PRIOR FILING DATE: 2000-08-03
; NUMBER OF SEQ ID NOS: 7059
; SOFTWARE: Patentin ver. 3.0
; SEQ ID NO 4772
; LENGTH: 513
; TYPE: PRT
; ORGANISM: Corynebacterium glutamicum
US-09-738-626-4772

  Query Match             4.6%;  Score 89;  DB 9;  Length 513;
  Best Local Similarity   21.2%;  Pred. No. 0.86;
  Matches   58;  Conservative  37;  Mismatches  86;  Indels  92;  Gaps  15;

Qy      86 IGEVVGGYLAHSLAVMTDAAHLLTDFASMLISLF-SLWMSSRPATKTMNFGWQRAEILGA 144
           :::: :: ::: :::: : : :::  :::::: :  ::::    :: : :   ::::
Db      87 VGDVFGGPLA------------LEGLIAFFLESVFLGLWI-------FGWGK---IPG 122

Qy     145 LVSVLSIWVV----TGVLVYLAV-----ERLISGDYEIDGGT-------MLITSGCAVAV 188
           ::: : : : :   : : :::  :   :: : : : :   :      :: :: ::
Db     123 WLHTASIWIVAIATNISAYFIIVANSFMQHPVGAEYNPETGRAELTDFWALLTNSTALAA 182

Qy     189 ---------------NIIMGLT--------LHQSGHGHSHGTTNQQEENPS--------VRA 219
                          :::::::       : :
Db     183 FPHAVAGGFLTAGTEVLGISGWWIIRAHRQAKKAESEIESKHSMHRPALWVGWWTTVVSS 242

Qy     220 AFIHVIGDFMQSMGVLVAAVILYFKPEYKY--VDPICTFV----FSILVLGT----TLT 268
           :: :          :     : :: :  : :  : :  ::   :::: :   :::
Db     243 VALFITGD------IQAKLMFVQQPMKMASAESLCETATDPNFSILTIGTHHNCDTVT 294

Qy     269 ILRDVLLV---LMEGTPKGVDFTAVRDLLSVE 298
           :: :: :     ::  :   :::  :: :: :
Db     295 HLIDVPFVLPFLAEGKFTGVILQGVNQLQAAAE 327
```

```
Qy     301 EALHSLHIWALTVAQPVLSVHIAIAQONTDAQAVLKTASSRLQG 343
           : : :: ::::  ::  : : : :    ::        :
Db     349 ----------LVMSQVLLSFGIALAL---VPLLIFTSDSKLMG 378
```

```
Db     304 VVGTLAGQVVMQGFIRFHIPLWVRRTVTMLPSFIVILM------GLDPTRI------- 348
           :::    :: ::         :::: :: : :            : :: ::
Qy     301 EALHSLHIWALTVAQPVLSVHIAIAQONTDAQAVLKTASSRLQG 343
```

Search completed: March 18, 2003, 15:46:16
Job time : 24 secs

```
OM protein - protein search, using sw model

                    GenCore version 5.1.4_p5_4578
                    Copyright (c) 1993 - 2003 Compugen Ltd.

Run on:             March 18, 2003, 15:39:05 ; Search time 19 Seconds
                                               (without alignments)
                                               1882.211 Million cell updates/sec

Title:              US-09-691-219-2
Perfect score:      1923
Sequence:           1 MEAKEKQHLLDARPAIRSYT...........QIEDYSEDMKDCQACQGPSD 372

Scoring table:      BLOSUM62
                    Gapop 10.0 , Gapext 0.5
Searched:           283224 seqs, 96134422 residues

Total number of hits satisfying chosen parameters:        283224

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing:    Minimum Match 0%
                    Maximum Match 100%
                    Listing first 120 summaries

Database :          PIR_73:*
                    1: pir1:*
                    2: pir2:*
                    3: pir3:*
                    4: pir4:*

    Pred. No. is the number of results predicted by chance to have a
    score greater than or equal to the score of the result being printed,
    and is derived by analysis of the total score distribution.
```

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	1493	77.6	359	1	S70632	zinc transporter 2
2	838.5	43.6	382	2	T24963	hypothetical prote
3	709.5	36.9	410	2	T26757	hypothetical prote
4	702	36.5	391	2	T26756	hypothetical prote
5	649.5	33.8	398	2	T20681	probable zinc tran
6	642	33.4	378	2	T49164	zinc transporter-1
7	599	31.1	334	2	T47986	zinc transporter-1
8	500.5	26.0	385	2	A84696	probable zinc tran
9	431	22.4	312	2	AE0593	cation-efflux syst
10	431	22.4	313	1	H64810	probable cation tr
11	430	22.4	313	2	D90726	probable transport
12	427	22.2	311	2	E85577	probable transport
13	421.5	21.9	321	2	G82752	cobalt-zinc-cadmiu
14	416.5	21.7	312	2	T35276	cobalt efflux pr
15	414.5	21.6	303	2	AG1396	cation transport p
16	410.5	21.3	303	2	AB1772	cation transport p
17	399.5	20.8	319	2	D89778	hypothetical prote
18	397	20.7	311	2	C69612	cation-efflux prote
19	393.5	20.5	301	2	AH3431	cobalt-zinc-cadmiu
20	391	20.3	325	2	G90008	hypothetical prote
21	390	20.3	325	2	T44365	cation-efflux prote
22	387	20.1	316	1	JC4701	cadmium, zinc, cob
23	380.5	19.8	312	2	AI0138	probable cation tr
24	374.5	19.5	361	2	F87286	cation efflux fami
25	372	19.3	299	2	C83595	probable cation ef
26	357	18.6	442	2	S56057	heavy metal ion re
27	351	18.3	325	2	A75437	cation efflux syst
28	346.5	18.0	308	2	E70392	cation efflux syst
29	337	17.5	439	2	S58327	cobalt accumulatio
30	336.5	17.5	303	2	H97212	co/Zn/Cd efflux sy
31	335	17.4	507	2	S54303	zinc transport pro
32	334.5	17.4	284	2	G72363	cation efflux syst
33	334	17.4	300	2	D84459	probable cation tr
34	333	17.3	300	2	AG2540	zinc transporter 2
35	332	17.3	503	2	S54302	cation efflux syst
36	315	16.4	304	2	H97467	cation transport p
37	314	16.3	387	2	H95216	probable heavy met
38	309	16.1	316	2	AC2686	probable zinc/cadm
39	296.5	15.4	295	2	G86825	probable cation tr
40	245.5	12.8	199	1	T43140	cation efflux syst
41	239	12.4	549	1	T38252	hypothetical prote
42	233	12.1	304	2	E81321	cation transport p
43	230.5	12.0	335	2	H71078	probable inner mem
44	230.5	12.0	338	2	D33830	cation efflux syst
45	229	11.9	299	2	T155506	hypothetical prote
46	226.5	11.8	281	2	H98080	cation-efflux syst
47	206	10.7	271	2	AF2161	probable cation tr
48	204.5	10.6	326	2	H90394	probable metal tra
49	190	9.9	740	2	D83483	probable membrane tra
50	186.5	9.7	240	2	D83493	hypothetical prote
51	184	9.6	292	2	S61568	probable cation-ef
52	173	9.0	310	2	T50628	cation efflux syst
53	172	8.9	164	2	C86944	hypothetical prote
54	170	8.8	290	2	A97035	disease resistance
55	164	8.5	290	2	S75583	conserved hypothet
56	162	8.4	298	2	T07759	conserved hypothet
57	158.5	8.2	449	2	B69772	probable membrane
58	158	8.2	300	2	F95119	cation efflux syst
59	157.5	8.2	317	2	B69119	conserved hypothet
60	156	8.1	483	2	B82045	hypothetical prote
61	156	8.1	288	2	AG1353	conserved hypothet
62	156	8.1	291	2	D84337	conserved hypothet
63	156	8.1	384	2	AH1746	conserved hypothet
64	154	8.0	428	2	AH1745	zinc resistance pr
65	149	7.7	302	2	D83150	conserved hypothet
66	147.5	7.7	289	2	B69791	hypothetical prote
67	145	7.5	299	2	T43145	cation efflux syst
68	143.5	7.5	291	2	E96974	conserved hypothet
69	143	7.4	298	2	AH1723	probable transport
70	142.5	7.4	356	2	C69023	cation efflux syst
71	142	7.4	403	2	C69781	hypothetical prote
72	139	7.2	289	2	C82411	C-terminal region
73	139	7.2	407	2	F72323	hypothetical prote
74	136	7.1	297	2	E96910	conserved hypothet
75	135.5	7.0	317	2	G70477	cation efflux syst
76	134.5	7.0	306	2	A12855	hypothetical 32.9K
77	134	6.9	292	2	F97632	hypothetical prote
78	133.5	6.9	385	2	H70941	probable transport
79	133	6.9	318	2	S55124	cation efflux syst
80	133	6.9	318	2	B98048	cation efflux syst
81	133	6.9	332	2	H83738	hypothetical prote
82	132.5	6.9	510	2	H83037	probable membrane
83	132	6.8	394	2	S40858	probable transport
84	131	6.8	307	2	H86080	cation efflux syst
85	130.5	6.8	300	2	H91233	hypothetical prote
86	130.5	6.8	300	2	F95180	probable transport
87	130.5	6.8	300	2	A64356	cation transport
88	130	6.8	394	2	A10942	probable transport
89	130	6.6	283	2	C81369	gabP 5'-region hyp
90	126	6.6	298	2	A69517	probable transmemb
91	123	6.4	295	2	T40838	cation efflux syst
92	122.5	6.4	283	2	G70176	probable transmemb
93	121	6.3	397	2	F83804	conserved hypothet
94	119	6.2	283	2	D86663	probable transport
95	118.5	6.2	451	2	D86664	conserved hypothet
96	118.5	6.2	425	2	AE0010	cation efflux syst
97	110.5	5.7	300	2	F86805	cation transporter
98	109.5	5.7	314	2	S74173	probable integral
99	109.5	5.7	300	2	H75109	cation efflux syst
100	108.5	5.6	4859	2	F86805	ryanodine receptor
101	107.5	5.6	314	2	S74173	cation transporter
102	107.5	5.6	283	2	H75109	cation efflux syst

103	107.5	5.6	302	2	AD3308	cobalt-zinc-cadmiu
104	107.5	5.6	398	2	B86298	protein F309.11 [i
105	107.5	5.6	434	2	E96826	hypothetical prote
106	107	5.6	251	2	T21114	hypothetical prote
107	107	5.6	251	2	T21114	hypothetical prote
108	107	5.6	397	2	AB2671	hypothetical prote
109	105	5.6	572	2	AC2110	cytochrome-c oxida
110	104.5	5.6	266	2	C81049	transporter, proba
111	104.5	5.4	515	2	T23089	hypothetical prote
112	103.5	5.4	680	2	T47777	hypothetical prote
113	103	5.4	301	2	C83703	hypothetical prote
114	102.5	5.4	377	2	T21751	hypothetical prote
115	102.5	5.4	379	2	S58449	ubiquinol-cytochro
115	102.5	5.3	410	2	S74951	Na+/H+-exchanging
116	102	5.3	509	2	T31797	hypothetical prote
117	101.5	5.3	474	2	S63658	NADH2 dehydrogenas
118	101.5	5.3	681	2	S75795	NADH2 dehydrogenas
119	101	5.3	306	2	G97480	NADH2 dehydrogenas
120	101	5.3	306	2	AG2698	hypothetical prote
						cation efflux syst

ALIGNMENTS

RESULT 1
S70632
zinc transporter ZnT-2 - rat
C;Species: Rattus norvegicus (Norway rat)
C;Date: 10-Sep-1999 #sequence_revision 10-Sep-1999 #text_change 21-Jul-2000
C;Accession: S70632
R;Palmiter, R.D.; Cole, T.B.; Findley, S.D.
A;Title: ZnT-2, a mammalian protein that confers resistance to zinc by facilitating vesi
A;Reference number: S70632; MUID:96203098; PMID:8617223
EMBO J. 15, 1784-1791, 1996
A;Status: preliminary
A;Molecule type: mRNA
A;Residues: 1-359 <PAL>
A;Cross-references: EMBL:U50927; NID:g1256377; PIDN:AAB02775.1; PID:g1256378
A;Start codon: CTG
A;Gene: ZnT-2
C;Genetics:
C;Superfamily: zinc transporter ZnT-2

Query Match 77.6%; Score 1493; DB 1; Length 359;
Best Local Similarity 80.1%; Pred. No. 4.5e-119;
Matches 290; Conservative 35; Mismatches 29; Indels 8; Gaps 2;

```
Qy  15 AIRSYTGSLWQEGAGWIPLPRPGDLDQAIELAAQSNHHCHAQKGPDSHCDPKKGKAQRQL  74
        | |:: |||| | :||:||||     ::||:: |:|:||||| ||| |:| | |||::||
Db   2 ASRSFPGALWKSEASRI----PPVNLPSVELAVQSNHIYCHAQKDSGSHPNSEKQRARRKL  57

Qy  75 YVASAICLLFMIGEVGGYLAHSLAVMTDAAHLLTDFASMLISLFSLWMSSRPATKTMNF 134
        |||||||:||||||:||||| ||| |||||||||||||||||||||||:|||| ||||
Db  58 YVASAICLVFMIGEIIGGYLAQSLAIMTDAAHLLTDFASMLISLFSLWVSSRPAIKTMNF 117

Qy 135 GWQRAEILGALVSVLSIWVVTGVLVYLAVERLISGCYEIDGGTMLITSGCAVAVNIIMGL 194
        |||||||||||:||||||||||||||||:||||| |||:| |||||||||||||||||||
Db 118 GWQRAEILGALLSVLSIWVVTGVLVYLAVQRLISGDYEIKGDTMLITSGCAVAVNIIMGL 177

Qy 195 TLHQSGHGHSHG----TTNQEENPSVRAAFIHVIGDFMQSMGVLVAAYILYFKPEYKYV 250
        :||||||||||     :: |||||||||||||:||: :||:|||||||:||||||||||
Db 178 ALHQSGHGHSHGSHEDSSQQOQNPSVRAAFIHVVGDLLQSVGVLVAAYIIYFKPEYKYV 237

Qy 251 DPICTFVFSIILVLGTTLTILRDVILVLMEGTPKGVDFTAVRDLLSVEGVEALHSLHIWA 310
        ||||||:|||||||||||||||||||||||||||||||| || : :|||:|||||||||
Db 238 DPICTFLFSIILVLGTTLTILRDVILVLMEGTPKGVDFTTVKNLLSVDGVEALHSLHIWA 297

Qy 311 LTVAQPVLSVHIAIAQNTDAQAVLKTPASSRLQGKFHFHTVTIQIEDYSEDMKDCQACQGP 370
        |||||||||||||||||.||||||| |  .|||||:|||.|||||: |||||:|| ||||
Db 298 LTVAQPVLSVHIAIAQNVDAQAVLKYARDRLQGKFNFHTMTIQIESYSEDMKSCQECQGP 357
```

Db 358 SE 359

RESULT 2
T24963
hypothetical protein T18D3.3 - Caenorhabditis elegans
C;Species: Caenorhabditis elegans
C;Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 18-Feb-2000
C;Accession: T24963
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-382 <WIL>
A;Cross-references: EMBL:Z68119; PIDN:CAA92193.1; GSPDB:GN00028; CESP:T18D3.3
C;Genetics:
A;Experimental source: clone T18D3
A;Gene: CESP:T18D3.3
A;Map position: X
A;Introns: 34/2; 72/1; 142/1; 193/2; 254/3; 290/1; 325/3

Query Match 43.6%; Score 838.5; DB 2; Length 382;
Best Local Similarity 48.6%; Pred. No. 1.8e-63;
Matches 161; Conservative 66; Mismatches 93; Indels 11; Gaps 3;

```
Qy  50 NHHCHAQ-KGPDSHCDPKKGKAQRQLYVASAICLLFMIGEVGGGYLAHSLAVMTDAAHLL 108
        | |||   ||||:| |:        ||||||||||:|:||||:|| ||| ||:|||||||
Db  53 NFHCHDEADSTDSH--DSNRRATRIILWLTVVLCLFFMVCEVIGGVIAGSLAIVTDAAHLL 110

Qy 109 TDFASMLISLFSLWMSSRPATKTMNFGWQRAEILGALVSVLSIWVVTGVLVYLAVERLIS 168
        |||||:|||||||:: :|| ::|  ||: ||::||| ||:|||:|||||||:||:.|:||
Db 111 TDFASVLISLFSLYIARRPPSQKMSFGFHRAEVLGAFSVFLIWIVTGVLVVLAIMRIVS 170

Qy 169 GDYEIDGGTMLITSGCAVAVNIIMGLTLHQSGHGHSHGTTNQ-------QEENPSVRAA 220
        |||| :||.|| |: | |||||:: : | ||||||||          |:|||.||||
Db 171 GDYEVEGGIMALTAALGVVVNLVMLALLYFGGHSHSHGGGSSHGHSHGGGNGDNINVRAA 230

Qy 221 FIHVIGDFMQSMGVLVAAYILYFKPEYKYVDPICTFVFSILVLGTTLTILRDVILVLMEG 280
        ||||:||::||:|||||| :|||:|  |:::|||||:||::||.|||.:||| | ||:||
Db 231 FIHVLGDLLQSLGVLVAALFIYFQPSWVIIDPICTLVFSVIVLCTTIYILRDAMIVLLEG 290

Qy 281 TPKGVDFTAVRDLLLSVEGVEALHSLHIWALTVAQPVLSVHIAIAQNTDAQAVLKTASSR 340
        .| ::||:.| . ||::::|: | :|: ||:::::| || ||:| .  :::::::||::|
Db 291 RPSNIDFAKVFSSLEDIEGVKKVHDLRTWSLIMDKIALSVHLEIDANSQSQSTLRETRKM 350

Qy 341 LQGKFHFHTVTIQIEDYSEDMKDCQACQGPS 371
        | |.:.:| :|||||::| : | |:|| :|
Db 351 LKQTYNVHEITIQIEEFGANRSDCGKCDFPT 381
```

RESULT 3
T26757
hypothetical protein Y39E4A.2b - Caenorhabditis elegans
C;Species: Caenorhabditis elegans
C;Date: 15-Oct-1999 #sequence_revision 15-Oct-1999
C;Accession: T26757
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-410 <WIL>
A;Cross-references: EMBL:AL021400; PIDN:CAA16328.1; GSPDB:GN00021; CESP:Y39E4A.2b
submitted to the EMBL Data Library, January 1998
R;Barlow, K.
A;Reference number: Z20260
A;Experimental source: clone Y39E4A
C;Genetics:
A;Gene: CESP:Y39E4A.2b
A;Map position: 3
A;Introns: 97/3; 362/1

Query Match 36.9%; Score 709.5; DB 2; Length 410;

```
Qy 371 LTVAQPVLSVHIAIAQNTDAQ 372
        |:    |
Db 371 SD 372
```

```
Qy   52  HCHAQKGPDSHCD----------------PKKGKAQRQLYVASAICLLFMIGEVVGGYI  94
         ||:|                        |::| ||| |||||:|||||||||||
Db   66  HSHTHNNDDSSSDCSGAGGGAHHKHSHSHDEKYQKGRAEKVLWAVAALSAVFVAAFVGGFW 125

Qy   95  AHSLAVMTDAAHLLTDFASMLISLFSLWMSSRPATKTMNFGWQRAEILGALVSVLSIWVV  154
         | |||:||||.|:|:|::|:|||:|::||::|||:||| ||||:||||||.||:.|:|:|
Db  126  AQSLAIMTDAGHHMLSDLLSFIISIFAIRCARLPASKRLSFGYERAEVLGALTSVIILWVL 185

Qy  155  TGVLVYLAVERLISGDYEIDGGTMLITSGCAVAVNIIMGLTLH--OSGHGHSHGTTN--  209
         |.|||:.|:.||:..:.||||| |:||.|.|.|.|||||| || ||||||||: :
Db  186  TVLVVVAIQRIVNNEHEVDADVMLITAGVGVLFNIVMGLVLHFGTGGHGHSSHG      245

Qy  210  --QOEENPSVRAAFIHVIGDFMQOSMGVLVAAYILYFKPEYKYVDPICTFVFSILVLGTTL 267
           |::|||||||||:|||||::| |:||:|||:||:.   .|:||||| :||::||:.|:|
Db  246  HAHDGKNVNVRAALIHVIGDLVQSIGVLIAALIIRF-TGWTLADPICTLFSIIVLFTTV    304

Qy  268  TILRDVILVLMEGTPKGVDFTAVRDLLLSVEGVEALHSLHIWALTVAQPVLSVHIAIAQN  327
         |::|| ::||||.||.|.|::|.:|| |:|||:.:|.||::|::..||||:|||:|:|::
Db  305  TVMRDIFVLMEATPSHYDLSDVKKALSALEGVRGVHDLHLWSIGMDKTAFSVHLALESP   364

Qy  328  TDAQAVLKTASSRLQGKFHFHTVTIQIEDYSEDMKDCQACQ 368
         . ..:|:||:|:|::| |:|:||..:|:|:.|:  .|.|||
Db  365  NRAMENVAEARSLIRRRFGVAVATVQVEPFDEKIDSCDTCQ 405
```

RESULT 4
T26756
hypothetical protein Y39E4A.2a - Caenorhabditis elegans
C;Species: Caenorhabditis elegans
A;Accession: T26756
C;Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 15-Oct-1999
A;Molecule type: DNA
A;Residues: 1-391 <WIL>
R;Barlow, K.
submitted to the EMBL Data Library, January 1998
A;Reference number: Z20260
A;Accession: T26756
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Cross-references: EMBL:AL021480; PIDN:CAA16327.1; GSPDB:GN00021; CESP:Y39E4A.2a
A;Experimental source: clone Y39E4A
C;Genetics:
A;Gene: CESP:Y39E4A.2a
A;Map position: 3
A;Introns: 22/1; 78/3; 343/1

Query Match   36.5%;   Score 702;   DB 2;   Length 391;
Best Local Similarity   40.6%;   Pred. No. 7e-52;
Matches 138;   Conservative 76;   Mismatches 102;   Indels 24;   Gaps 5;

```
Qy   52  HCH--AQKGPDSHCDPKKG-----------------KAQRQLYVASAICLLFMIGEVVGGYIA  95
         |||   |:||:||                       |||||||||||||||||||||||||.
Db   48  HCHYWDENDDDMVARVERGGSTDSASSREDTGRRAEKVLWAVAALSAVFIAAEFVGGEWA    107

Qy   96  HSLAVMTDAAHLLTDFASMLISLFSLWMSSRPATKTMNFGWQRAEILGALVSVLSIWVVT  155
         ||||:||||:||:||:|::||:||:.:.||:|:.:|.|.|.||.||:|||.|||.:.|:|
Db  108  QSLAIMTDAGHMLSDLLSFIISIFAIRCARLPASKRLSFGYERAEVLGALTSVIILWVLT 167

Qy  156  GVLVYLAVERLISGDYEIDGGTMLITSGCAVAVNIIMGLTLH--OSGHGHSHGTTN---  209
         .||:.|||.||:|||||:|||| |:||.|.|.|.|||||| ||  ||||||||:
Db  168  TVLVVVAIQRIVNNEHEVDADVMLITAGVGVLFNIVMGLVLHFGTGGHGHSSHGH      227

Qy  210  -QOEENPSVRAAFFHVICIDFMQOSMGVLVAAYILYFKPEYKYVDPICTFVFSILVLGTTLT 268
           |::|||||||||:|||:||::| |:|||:|||:||:.   .:|:||||| :||::||:.|||
Db  228  AHDGKNVNVRAALIHVIGDLVQSIGVLIAALIIRF-TGWTLADPICTLFSIIVLFTTV    286

Qy  269  ILRDVILVLMEGTPKGVDFTAVRDLLLSVEGVEALHSLHIWALTVAQPVLSVHIAIAQNT 328
         |:|| ::||||.||.|.|::|.:||| :|||:.:|.||::|::..||||:|||:|:|::.
Db  287  VMRDIFVLMEATPSHYDLSDVKKALSALEGVKGVHDLHLWSIGMDKTAFSVHLALESPN  346

Qy  329  DAQAVLKTASSRLQGKFHFHTVTIQIEDYSEDMKDCQACQ 368
         . ..:|:||:|:|::| |:|:||..:|:|:.|:  .|.|||
Db  347  RAMENVAEARSLIRRRFGVAVATVQVEPFDEKIDSCDTCQ 386
```

RESULT 5
T02681
probable zinc transporter At2g46800 [imported] - Arabidopsis thaliana
N;Alternate names: hypothetical protein F19D11.8
C;Species: Arabidopsis thaliana (mouse-ear cress)
C;Date: 24-Mar-1999 #sequence_revision 24-Mar-1999 #text_change 16-Feb-2001
A;Accession: T02681; D84907
R;Rounsley, S.D.; Lin, X.; Kaul, S.; Shea, T.P.; Fujii, C.Y.; Mason, T.M.; Shen, M.;
submitted to the EMBL Data Library, September 1998
A;Description: Arabidopsis thaliana chromosome II BAC F19D11 genomic sequence.
A;Reference number: 214698
A;Accession: T02681
A;Molecule type: DNA
A;Residues: 1-398 <ROU>
R;Lin, X.; Kaul, S.; Rounsley, S.D.; Shea, T.P.; Benito, M.I.; Town, C.D.; Fujii, C.Y
A;Experimental source: cultivar Columbia
M.; Koo, H.; Moffat, K.S.; Cronin, L.A.; Shen, M.; Vanaken, S.E.; Umayam, L.; Tallon,
euss, D.; Nierman, W.C.; White, O.; Eisen, J.A.; Salzberg, S.L.; Fraser, C.M.; Venter
Nature 402, 761-768, 1999
A;Title: Sequence and analysis of chromosome 2 of the plant Arabidopsis thaliana.
A;Reference number: A84420; MUID:20083487; PMID:10617197
A;Accession: D84907
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-398 <STO>
A;Cross-references: GB:AE002093; NID:g3510254; PIDN:AAC33498.1; GSPDB:GN00139
C;Genetics:
A;Gene: At2g46800; F19D11.8
A;Map position: 2

Query Match   33.8%;   Score 649.5;   DB 2;   Length 398;
Best Local Similarity   37.9%;   Pred. No. 2e-47;
Matches 134;   Conservative 71;   Mismatches 88;   Indels 61;   Gaps 4;

```
Qy   60  DSHCDPKKGKAQRQLYVASAICLLFMIGEVVGGYLAHSLAVMTDAAHLLTDFASMLISLF 119
         |:|   ||||||||||||||||||||||||||||:|||||:||||:||:||:|::||:||
Db   47  DAH--ERSASMRKLCIAVVLCLVFMSVEVVGGIKANSLAILTDAAHLLSDVAAFAISLF   103

Qy  120  SLWMSSRPATKTMNFGWQRAEILGALVSVLSIWVVTGVLVYLAVERLISGDYEIDGGTMT 179
         |||:|||:|::|::    :|:.|:.|.:|:.::.:::.|:|.||||||||||||||||| L
Db  104  SLWAAGWEATPRQTYGFFRELGALVSIQLIWLLTGILVYEAIIRIVTERSEVNGFLMF   163

Qy  180  ITSGCAVAVNIIMGLTLH-----OSGHGHSHG----------------TTN---------  209
         ||||.|.|.|||||| ||      ||||||||                |||
Db  164  LVAAFGLVVNIIMAVLLGHDHGHSHGHGHDHHNHSHGVTVTHHHHNHHNDHEHGHSHG    223

Qy  210  -----------------QOEENPSVRAAFIHVICIDFMQOSMGVLVAAYIL 241
                          :|::||||||||||:|||:||::|:|:||:|||:
Db  224  HGEDKHHAHGDVTEQLLDKSKTQVAAKEKRRNINLQGAYLHVLGDSIQSVGVMIGGAII 283

Qy  242  YFKPEYKYVDPICTFVFSILVLGTTLTILRDVILVLMEGTPKGVDFTAVRDLLLSVEGVE 301
         ||:   :.|:||||| :||::||||:|||:: |::|||.||.|.|::|.:||| :|||:.
Db  284  WYNPEWKIVDLICTLAFSVIVLGTTINMIRNILEVLMESTPREIDATKLEKGLLEMEEVV 343

Qy  302  ALHSLHIWALTVAQPVLSVHIAIAQNTDAQAVLKTASSRLQGKFHFHTVTIQIE 355
         .:|.||:|::..||||:|||:|:|::.. ..:|:||:|:|::| |:|:||..:|
Db  344  AVHELHIWAITVGKVLLACHVNIRPEADADMVLNKVIDYIRREYNISHVTIQIE 397
```

RESULT 6
T49164
zinc transporter-like protein - Arabidopsis thaliana
N;Alternate names: protein T20N10.160
C;Species: Arabidopsis thaliana (mouse-ear cress)
C;Date: 02-Jun-2000 #sequence_revision 02-Jun-2000 #text_change 02-Jun-2000
A;Accession: T49164
R;D'Angelo, M.; Vezzi, A.; Modesto, D.; Pigazzi, M.; Valle, G.; Mewes, H.W.; Rudd, S.

submitted to the Protein Sequence Database, April 2000
A:Reference number: Z25017
A:Accession: T49164
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-378 <DAN>
A:Cross-references: EMBL:AL353032; GSPDB:GN00061; ATSP:T20N10.160
A:Experimental source: cultivar Columbia; BAC clone T20N10
C:Genetics:
A:Gene: ATSP:T20N10.160
A:Map position: 3

Query Match   33.4%;   Score 642;   DB 2;   Length 378;
Best Local Similarity   39.3%;   Pred. No. 8.3e-47;
Matches   128;   Conservative   69;   Mismatches   93;   Indels   36;   Gaps   2;

```
Qy  66 KKGKAQRQLYASAICLLFMIGEVVGGYLAHSLAVMTDAAHLLTDFASMLISLFSLWMSS 125
        |:|:||  ||| | ::| ||||  |:|:| ||||:||||||||||||||||||| |
Db  51 ERAAASMRKLLIAVLLCAIFIVVEVVGGIKANSLAILTDAAHLLSDVAAFAISLFSLWASG 110

Qy 126 RPATKTMNFGWQRAEILGALVSVLSIWVVTGVLVYLAVAVERLISGDYEIDGGTMLITSGCA 185
        || ||||  | |:|||| ||||||||:::|| |||||::||||||||||||||| |
Db 111 WKANPQQSYGFFRIEILGALVYSIQMIWLLAGTLIVYEAIVRLNNGSGEVEGSLMFAVSAVG 170

Qy 186 VAVNIIMGLTL---------------HQSGHGHSHG----------------TTN 209
        ||||| ||| | : |||||| |
Db 171 LLVNIAMAILLGHDHGHGHSHDNGHGHSHDHGHGIAATEHHHHDSGGHDESQLSDVLIEQ 230

Qy 210 QQEENPSVRAAFIHVIGDFMQSMGVLVAAYILYFKPEYKYVDPICTFVFSILVLGTTLTI 269
        :::: | | ||:||:||| ||:||| ||| :|||||:|||||:||||:|||||||: :
Db 231 KKQRNVNIQAYLHVLGDSIQSVGVMIGGAIIWYKPEWKILDLICTLVFSVILVGTTIGM 290

Qy 270 LRDVILVLMEGTPKGVDFTAVRDLLLSVEGVEALHSLHIWALTVAQPVLSVHIAIAQNTD 329
        || ||:|||  || :| :: | || :||:||| ||:||||||||: ||:|| |: |
Db 291 LRNILEVLMESTPREIDPTMLEKGVCEIEEVVAVHELHIWAITVGKLLLACHVKIRPEAE 350

Qy 330 AQAVLKTASSRLQGKFHFHTVTIQIE 355
        | ||| | |||||||
Db 351 ADMVLDKIIDYIKREHNISHVTIQIE 376
```

RESULT 7
T47986
zinc transporter-like protein - Arabidopsis thaliana
N:Alternate names: protein F21F14.110
C:Species: Arabidopsis thaliana (mouse-ear cress)
C:Date: 20-Apr-2000 #sequence_revision 20-Apr-2000 #text_change 27-Nov-2001
C:Accession: T47986
R:Choisne, N.; Robert, C.; Brottier, P.; Wincker, P.; Cattolico, L.; Artiguenave, F.; Sa
submitted to the Protein Sequence Database, February 2000
A:Reference number: Z24481
A:Accession: T47986
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-334 <CHO>
A:Cross-references: EMBL:AL138642
A:Experimental source: cultivar Columbia; BAC clone F21F14
C:Genetics:
A:Map position: 3
A:Note: F21F14.110
C:Superfamily: zinc transporter ZnT-2

Query Match   31.1%;   Score 599;   DB 2;   Length 334;
Best Local Similarity   40.0%;   Pred. No. 3.2e-43;
Matches   126;   Conservative   64;   Mismatches   105;   Indels   20;   Gaps   4;

```
Qy  61 SHCDPKKGKA-QRQLYASAICLLFMIGEVVGGYLAHSLAVMTDAAHLLTDFASMLISLF 119
         | | ||:| :|:| |:| ||||| ||||  |:|:|||:|||||||||:| ||:|:|
Db  19 SSSDAKKRAASMRKLCFVVVLCLLFMSIEVVCGIKANSLAILADAAHLLTDVGAFAISML 78

Qy 120 SLWMSSRPATKTMNFGWQRAEILGALVSVLSIWVVTGVLVYLAVAVERLI-SGDYEIDGGTM 178
        |||:||  | | |||||| ||||:||||:: ||::| ||| ||:|  :||:
Db  79 SLWASSWEANPQSYGFFRIEILGTLVSIQLIWLLIGILVEAVTRLVQETNDDVDGFFM 138
```

```
Qy 179 LITSGCAVAVNIIMGLTL-HQSGHGHSGTT-------------------NQQEENPSVRAA 220
        |||| |||| : || : : :|||||||: ::|:|
Db 139 VLVAAFGLVVNIIMTIVLGHDHGHDGHGHSHDGHGHSYGERAEQLLEKSKEIRNIVQGA 198

Qy 221 FIHVIGDFMQSMGVLVAAYILYFKPEYKYVDPICTFVFSILVLGTTLTILTILRDVILVLMEG 280
        :||:|||::||:||:: | :::||:|: ||||:|| |:||||| :|||| |:||:|||
Db 199 YLHVLGDLIQSIGVMIGGGMIWNPKWKVIDLICTLFFSVILVGTTIKMLRSILEVLMES 258

Qy 281 TPKGVDFTAVRDLLLSVEGVEALHSLHIWALTVAQPVLSVHIAIAQNTDAQAVLKTASSR 340
        || :| | |: :: :||:|::|||||||:|||: | |||: :: ::
Db 259 TPREIDARQLEKGLMEIEEVVDVHELHIWAITVGKALFSCHVKVRPEAGDEMVLNKVIDY 318

Qy 341 LQGKFHFHTVTIQIE 355
         ::| ::|||||||
Db 319 IWREYRISHVTIQIE 333
```

RESULT 8
A84696
probable zinc transporter [imported] - Arabidopsis thaliana
C:Species: Arabidopsis thaliana (mouse-ear cress)
C:Date: 02-Feb-2001 #sequence_revision 02-Feb-2001 #text_change 02-Feb-2001
C:Accession: A84696
R:Lin, X.; Kaul, S.; Rounsley, S.D.; Shea, T.P.; Benito, M.I.; Town, C.D.; Fujii, C.Y.
M.; Koo, H.; Moffat, K.S.; Cronin, L.A.; Shen, M.; VanAken, S.E.; Umayam, L.; Tallon,
euss, D.; Nierman, W.C.; White, O.; Eisen, J.A.; Salzberg, S.L.; Fraser, C.M.; Venter
Nature 402, 761-768, 1999
A:Title: Sequence and analysis of chromosome 2 of the plant Arabidopsis thaliana.
A:Reference number: A84420; MUID:20083487; PMID:10617197
A:Accession: A84696
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-385 <STO>
A:Cross-references: GB:AE002093; NID:g3980394; PIDN:AAC95197.1; GSPDB:GN00139
C:Genetics:
A:Gene: At2g29410
A:Map position: 2

Query Match   26.0%;   Score 500.5;   DB 2;   Length 385;
Best Local Similarity   29.8%;   Pred. No. 8.7e-35;
Matches   114;   Conservative   81;   Mismatches   130;   Indels   57;   Gaps   5;

```
Qy  11 DARPAIRSYTGSLWQEGAGWIPLPRPGILDLQAIELAAQSNHHCHAQKGPDSHCDPKKGKA 70
        | | ::: : ::| | | ::|:| | :: | |:|
Db  22 DDEEEMESPSPSKTEENLGVVPL------------SCAFTRQEHCVSE------TKEREES 64

Qy  71 QRQLYASAICLLFMIGEVVGGYLAHSLAVMTDAAHLLTDFASMLISLFSLWMSSRPATK 130
        :| | |:| :::::|:|:||| |||||||||||:| |::|:::| |:|| | :
Db  65 TRRISSLIFIYLIVMSVQIVGGFKANSLAVMTDAAHLLSDVAGLCVSLLAIKVSSWEANP 124

Qy 131 TMNFGWQRAEILGALVSVLSIWVVTGVLVYLAVAVERLISGDYEIDGGTMLITSGCAVAVNI 190
        ::.|| | || :|:||: |||:||::: | :|||: ::|:| |::|| |
Db 125 RNSFGFKRLEVLAAFLSVQLIWLVSGVIIHEAIQRLLSRSREVNGEIMFGISARFGFMNL 184

Qy 191 IMGLTLHQSGHGHSH------------------GTTNQQEENP----------------- 215
        :| || : |||:|| |::.::|:|
Db 185 VMVLWL---GHNHSHHHHHDHHHHHHHKEVVAEEEEEMNPLKGERSSSSKEM 241

Qy 216 --SVRAAFIHVIGDFMQSMGVLVAAYILYFKPEYKYVDPICTFVFSILVLGTTLTILTILRDV 273
          || | :|| :| :||:||:: | ::::||:|: ||||:| |:| | || :
Db 242 NINIQGAYLHAMADMIQSLGVMIGGGIIWVKPKNVLVDLICTLVFSAFALAATLPILKNI 301

Qy 274 ILVLMEGTPKGVDFTAVRDLLLSVEGVEALHSLHIWALTVAQPVLSVHIAIAQNTDAQAV 333
        |:|| | :| :| :: | ::||:|::| ||||:| |: ||:| | :
Db 302 FGILMERVPRDMDIEKLERGLKRIDGVKIVYDLHVWEITVGRIVLSCHILPEGASPKEI 361

Qy 334 LKTASSRLQGKFHFHTVTIQIE 355
         :|: | :: | :|:|:||:|
Db 362 ITGVRNFCRKSYGIYHATVQVE 383
```

RESULT 9

AE0593

probable cation transport protein ybgR [imported] - Salmonella enterica subsp. enterica
C;Species: Salmonella enterica subsp. enterica serovar Typhi
F;Note: this species has also been called Salmonella typhi
A;Title: Complete genome sequence of a multiple drug resistant Salmonella enterica serov
A;Reference number: AE0502; PMID:11677608
C;Date: 09-Nov-2001 #sequence_revision 09-Nov-2001 #text_change 27-Nov-2001
C;Accession: AE0593
R;Parkhill, J.; Dougan, G.; James, K.D.; Thomson, N.R.; Pickard, D.; Wain, J.; Churcher,
th; T.; Connerton, P.; Cronin, A.; Davis, P.; Davies, R.M.; Dowd, L.; White, N.; Farrar,
R;Parkhill, J.; Mungall, K.; Skelton, J.; Stevens, K.;
, S.; Moule, S.; O'Gaora, P.
Nature 413, 848-852, 2001
A;Authors: Parry, C.; Quail, M.; Rutherford, K.; Simmonds, M.; Skelton, J.; Stevens, K.;
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-312 <PAR>
A;Cross-references: GB:AL513382; PIDN:CAD05215.1; PID:gi16501985; GSPDB:GN00176
A;Genetics:
A;Gene: ybgR
C;Superfamily: zinc transporter ZnT-2

Query Match   22.4%;   Score 431; DB 2; Length 312;
Best Local Similarity   33.3%; Pred. No. 5.4e-29;
Matches 108; Conservative 66; Mismatches 130; Indels 20; Gaps 8;

Qy   49 SNNHCHAQKGPDSHCDPKKGKAQRQLYVASAICLLFMIGEVVGGYLAHSLAVMTDAAHLL 108
         |||| : |: :|||| | : : :: | ||| ||: ||| | |: ||: |||||
Db   2 AHSHSHA---DSHL-PKDNNARRLLF-AFIVTAGMLEVVGGILSGSLALLADAGHML 55

Qy 109 TDFASMLISLFSLWMSSRPATKTMNFGWQRAEILGALVSVLSIWVVTGVLVYLAVERLIS 168
         ||: |:|||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   56 TDAALLFALLVVQFSRRPPTVRHTFGWLRLTTLAAFVNAIALVVITILIIVWEAIERFYT 115

Qy 169 GDYEIDGGTMLITSGCAVAVNIIMGLTLHQSGHGHSHGTTNQQEENPSVRAAFTHVIGDF 228
         |||:||||||||||||||||||||||||||||||||||||||||||||||||||||:||
Db 116 -PRPVAGNLMMVIAVAGLLANLFAFWLHR------------GSDEKNLNVRAAALHVMGDL 164

Qy 229 MQSMGVLVAAYILYFKPEYKYVDPICTFVFSILVLGTTLTILRDVILVLMEGTPKGVDFT 288
         |||||||||||||||||||||||||||:||||||||||||||||||||||||||||||:
Db 165 LGSVGAIVAALIIIW-TGWTPADPILSILVSIVLVLRSAWRLLKDSVNELLEGAPVSLDIN 223

Qy 289 AV-RDLLLSVEGVEALHSLHIWALTVAQPVLSVHIAIAQNTDAQAVLKTASSRLQGKFHF 347
         || ||||||||||||| :|:||||::||||| | | | ||||:|:|:|||||||::|:
Db 224 ALQRHLSREIPEVRNVHVHVHVHVHVM-MVGEKPVMTLHAQVIPPHDHDALLERIQDFLMHEYHI 282

Qy 348 HTVTIQIEDYSEDMKDCQACQGPS 371
         ||| ||
Db 283 AHATIQMEYQMCHGPDCHLNQTPS 306

RESULT 10
H64810
ybgR protein - Escherichia coli (strain K-12)
C;Species: Escherichia coli
C;Date: 10-Sep-1999 #sequence_revision 10-Sep-1999 #text_change 01-Mar-2002
C;Accession: H64810
R;Blattner, F.R.; Plunkett III, G.; Bloch, C.A.; Perna, N.T.; Burland, V.; Riley, M.; Co
A;Rose, D.J.; Mau, B.; Shao, Y.
Science 277, 1453-1462, 1997
A;Title: The complete genome sequence of Escherichia coli K-12.
A;Reference number: A64720; MUID:97426617; PMID:9278503
A;Accession: H64810
A;Status: nucleic acid sequence not shown; translation not shown
A;Molecule type: DNA
A;Residues: 1-313 <BLAT>
A;Cross-references: GB:AE000177; GB:U00096; NID:gi1786955; PIDN:AAC73839.1; PID:gi1786966;
C;Experimental source: strain K-12, substrain MG1655
C;Genetics:
C;Gene: ybgR
C;Superfamily: zinc transporter ZnT-2
C;Keywords: transmembrane protein
F;21-37/Domain: transmembrane protein #status predicted <TM1>

F;54-70/Domain: transmembrane #status predicted <TM2>
F;90-106/Domain: transmembrane #status predicted <TM3>
F;126-142/Domain: transmembrane #status predicted <TM4>
F;164-180/Domain: transmembrane #status predicted <TM5>
F;188-204/Domain: transmembrane #status predicted <TM6>

Query Match   22.4%;   Score 431; DB 1; Length 313;
Best Local Similarity   32.7%; Pred. No. 5.5e-29;
Matches 104; Conservative 69; Mismatches 121; Indels 24; Gaps 9;

Qy   54 HAQKGPDSHCDPKKGKAQRQLYVASAICLLFMIGEVVGGYLAHSLAVMTDAAHLLTDFAS 113
         |: |||| | : : :: | ||| ||: ||| | |: ||: ||||:|||| : ||:
Db   3 HSHSHSSHL-PEDNNARRLLY-AFGVTAGFMLVEVVGGFLSGSLALLADAGHMLTDTAA 60

Qy 114 MLISLFSLWMSSRPATKTMNFGWQRAEILGALVSVLSIWVVTGVLVYLAVERLISGDYEI 173
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   61 LLFALLAVQFSRRPPTIRHFGWLRLTTLAAFVNAIALVVITILIIVWEAIERFRT-PRPV 119

Qy 174 DGGTMLITSGCAVAVNIIMGLTLHQSGHGHSHGTTNQQEENPSVRAAFTHVIGDFMQSMG 233
         |||||||||||||||||||||||||||||||||||||||||||||||||||:|||:|:|
Db 120 EGGMMNAIAVAGLLANILSFWLLH------HGS--EEKNLNVRAAALHVGDLLGSVG 169

Qy 234 VLVAAYILYFKPEYKYVDPICTFVFSILVLGTTLTILRDVILVLMEGTPKGVDFTAV-RD 292
         ||||:||||||||||||||||||:||||||||||||||||||||||||||||||: ||
Db 170 AIIAALIIIW-TGWTPADPILSILVSIVLVLRSAWRLLKDSVNELLEGAPVSLDIAELKRR 228

Qy 293 LLLSVEGVEALHSLHIWALTVAQPVLSVHIAIAQNTDAQAVLKTASSRLQGKFHFHTVTI 352
         |||||||||||| :|:||||::||||| | | | ||||:|:|:|||||||:|:| |
Db 229 MCREIPEVRNVHVHVHVHVM-MVGEKPVMTLHQVIPPHDHDALLDQOHYLMDHYQIEHATI 287

Qy 353 QIEDYSEDMKDCQACQGP 370
         || || : ||||||||||
Db 288 QME--------YQPCHGP 297

RESULT 11
D90726
probable transport system permease protein ECs0780 [imported] - Escherichia coli (str
C;Species: Escherichia coli
C;Date: 18-Jul-2001 #sequence_revision 18-Jul-2001 #text_change 03-Aug-2001
C;Accession: D90726
R;Hayashi, T.; Makino, K.; Ohnishi, M.; Kurokawa, K.; Ishii, K.; Yokoyama, K.; Han, C
gasawara, N.; Yasunaga, T.; Kuhara, S.; Shiba, T.; Hattori, M.; Shinagawa, H.
DNA Res. 8, 11-22, 2001
A;Title: Complete genome sequence of enterohemorrhagic Escherichia coli O157:H7 and g
A;Reference number: A99629; MUID:21156231; PMID:11258796
A;Accession: D90726
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-313 <HAY>
A;Cross-references: GB:BA000007; PIDN:BAB34203.1; PID:gi13360239; GSPDB:GN00154
A;Experimental source: strain O157:H7, substrain RIMD 0509952
A;Genetics:
A;Gene: ECs0780
A;Superfamily: zinc transporter ZnT-2

Query Match   22.4%;   Score 430; DB 2; Length 313;
Best Local Similarity   32.4%; Pred. No. 6.6e-29;
Matches 103; Conservative 70; Mismatches 121; Indels 24; Gaps 9;

Qy   54 HAQKGPDSHCDPKKGKAQRQLYVASAICLLFMIGEVVGGYLAHSLAVMTDAAHLLTDFAS 113
         |: |||| | : : :: | ||| ||: ||| | |: ||: ||||:|||| : ||:
Db   3 HSHSHTSSHL-PEDNNARRLLY-AFGVTAGFMLVEVIGGFLSGSLALLADAGHMLTDTAA 60

Qy 114 MLISLFSLWMSSRPATKTMNFGWQRAEILGALVSVLSIWVVTGVLVYLAVERLISGDYEI 173
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   61 LLFALLAVQFSRRPPTIRHFGWLRLTTLAAFVNAIALVVITILIIVWEAIERFRT-PRPV 119

Qy 174 DGGTMLITSGCAVAVNIIMGLTLHQSGHGHSHGTTNQQEENPSVRAAFTHVIGDFMQSMG 233
         |||||||||||||||||||||||||||||||||||||||||||||||||||:|||:|:|
Db 120 EGGMMNAIAVAGLLANILSFWLLH------HGS--EEKNLNVRAAALHVGDLLGSVG 169

Qy 234 VLVAAYILYFKPEYKYVDPICTFVFSILVLGTTLTILRDVILVLMEGTPKGVDFTAV-RD 292

```
Db  170  AIIAALIIIW-TGWTPADPILSILVLSVHIAIAQNTDAQAVLKTASSRLQKFHFHTVTI 228
          :::||:: |   ::|||| |:::  |||:|| : :: :  |:::||| |:|||:|:|
Qy  293  LLLSVEGVEALHSLHIWALTVAQPVLSVHIAIAQNTDAQAVLKTASSRLQKFHFHTVTI 352

Db  229  MCREIPEVRNVHVHVM-MVGEKPVMTLHVQVIPPHDHDALLDQIHQYLMDHYQIEHATI 287
Qy  353  QIEDYSEDMKDCQACGQP 370
          |:|  | ||
Db  288  QME------YOPCHGP 297
```

RESULT 12
E85577
probable transport system permease protein yBgR [imported] - Escherichia coli (strain O1...
C;Species: Escherichia coli
C;Date: 16-Feb-2001 #sequence_revision 16-Feb-2001 #text_change 14-Sep-2001
A;Accession: E85577
A;Experimental source: strain O157:H7, substrain EDL933
C;Genetics:
C;Gene: yBgR
C;Superfamily: zinc transporter ZnT-2

R;Perna, N.T.; Plunkett III, G.; Burland, V.; Mau, B.; Glasner, J.D.; Rose, D.J.; Mayhew...
ller, L.; Grotbeck, E.J.; Davis, N.W.; Lim, A.; Dimalanta, E.; Potamousis, K.; Apodaca,...
Nature 409, 529-533, 2001
A;Reference number: A85480; MUID:21074935; PMID:11206551
A;Title: Genome sequence of enterohemorrhagic Escherichia coli O157:H7.
A;Accession: E85577
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-311 <STO>
A;Cross-references: GB:AE005174; NID:g12513681; PIDN:AAG55081.1; GSPDB:GN00145; UWGP:209...

Query Match 22.2%; Score 427; DB 2; Length 311;
Best Local Similarity 32.5%; Pred. No. 1.2e-28;
Matches 104; Conservative 69; Mismatches 119; Indels 28; Gaps 10;

```
Qy   52  HCHAQKGPDSHCDPKKGKAQRQLYVASAICLLFMIGEVVGGYLAHSLAVMTDAAHLLTDF 111
Db    3  HSHT----SSHL-PEDNNARRLLY-AFGVTAGFMLVEVIGGFLSGSLLALLADAGHMLTDT  56

Qy  112  ASMLISLFSLWMSSRPATKTMNFGWQRAEILGALVSVLSIWVVTGIVLVYLAVERLISGDY 171
Db   57  AALLFALLAVQFSRRPTIRHTFGWLRLTTLAAFVNAIALVVITILIVWEAIERFFrt-PR 115

Qy  172  EIDGGTMLITSGCAVAVNIIMGLTLHQSGHGHSHGTTNQQEENPSVRAAFIHVIGDEMQS 231
Db  116  PVEGGMMMAIAVAGLLANILSFWLLH------HGS--EEKNLNYRAAALHVLGDLLGS 165

Qy  232  MGVLVAAYILYFKPEYKYVDPICTFVFSILLVGTTLTILRDVILVLMEGTPKGVDFTAV- 290
Db  166  VGAIIAALIIW-TGWTPADPILSILVLSVHIAIAQNTDAQAVLKTASSRLQKFHFHTVTV 224

Qy  291  RDLLLSVEGVEALHSLHIWALTVAQPVLSVHIAIAQNTDAQAVLKTASSRLQKFHFHTV 350
Db  225  RRMCREIPEVRNVHVHVM-MVGEKPVMTLHVQVIPPHDHDALLDQIHYLMDHYQIEHA 283

Qy  351  TIQIEDYSEDMKDCQACGQP 370
Db  284  TIQME------YOPCHGP 295
```

A;Accession: G82752
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-321 <SIM>
A;Cross-references: GB:AE003926; GB:AE003849; NID:g9105771; PIDN:AAP83676.1; GSPDB:GN...
A;Experimental source: strain 9a5c
R;Simpson, A.J.G.; Reinach, F.C.; Arruda, P.; Abreu, F.A.; Acencio, M.; Alvarenga, R....
Briones, M.R.S.; Bueno, M.R.P.; Camargo, A.A.; Camargo, L.E.A.; Carraro, D.M.; Carrer...
as-Neto, E.; Docena, C.; El-Dorry, H.; Facincani, A.P.; Ferreira, A.J.S.
submitted to GenBank, June 2000
A;Authors: Ferreira, V.C.A.; Ferro, J.A.; Fraga, J.S.; Franca, S.C.; Franco, M.C.; Fr...
chado, M.A.; Junqueira, M.L.; Kemper, E.L.; Kitajima, J.P.; Krieger, J.E.; Kuramae, E.E.; La...
, F.G.; Nunes, L.R.; Oliveira, M.C.; Madeira, H.M.F.; Marino, C.L.; Marques, M.V.; Martins
Rodrigues, V.; Rosa, A.J. de M.; de Rosa Jr., V.E.; de Sa, R.G.; Santelli, R.V.; Sawa...
A;Authors: da Silva, A.C.R.; da Silva, F.R.; da Silva Jr., W.A.; da Silv...
M.; Tsuhako, M.H.; Vallada, H.; Van Sluys, M.A.; Verjovski-Almeida, S.; Vettore, A.L.
A;Reference number: A59328
C;Contents: annotation
C;Genetics:
A;Gene: XF0866
C;Superfamily: zinc transporter ZnT-2

Query Match 21.9%; Score 421.5; DB 2; Length 321;
Best Local Similarity 32.3%; Pred. No. 3.6e-28;
Matches 98; Conservative 64; Mismatches 126; Indels 15; Gaps 5;

```
Qy   53  CHAQKGPDSHCDPKKGKAQRQLYVASAICLLFMIGEVVGGYLAHSLAVMTDAAHLLTDFA 112
Db   20  CH---MGHDHTHVPKGMHHERPLWWALGLTASFLVAEIIGAVLTRSLALLSDAVHMATDVF  77

Qy  113  SMLISLFSLWMSSRPATKTMNFGWQRAEILGALVSVLSIWVVTGVLVLYLAVERLISGDYE 172
Db   78  ALMIALAVRLSRRPPDVARRTYGYARLEAFGALVNGVLLFGVGGYILWEAVQRL-RAPQE 136

Qy  173  IDGGTMLITSGCAVAVNIIMGLTLHQSGHGHSHGTTNQQEENPSVRAAFIHVIGDEMQSM 232
Db  137  IVSSGMLVIAVLGIVINLIVMRLLH-AGRG-------ENLSMKGAYLEVWSDMLGSV 185

Qy  233  GVLVAAYILYFKPEYKYVDPICTFVFSILLVGTTLTILRDVILVLMEGTPKGVDFTAVRD 292
Db  186  AVIVAIVIYTGWY-WVDPVLAVLIGLWVLPRTWVLLGEAVNVLLEGVPKGFELLPIRD 244

Qy  293  LLLSVEGVEALHSLHIWALTVAQPVLSVHIAIAQNTDAQAVLKTASSRLQKFHFHTVTI 352
Db  245  ALSHHPGVANVHDVHLWALGSRMPLLTAHVVALQGTAPDRLRRMLQELLHERFSIEHVTL 304

Qy  353  QIE 355
Db  305  QIE 307
```

RESULT 13
G82752
cobalt-zinc-cadmium resistance protein XF0866 [imported] - Xylella fastidiosa
C;Species: Xylella fastidiosa
C;Date: 18-Aug-2000 #sequence_revision 20-Aug-2000 #text_change 02-Sep-2000
R;Accession: G82752
R;anonymous,
R;Title: The Xylella fastidiosa Consortium of the Organization for Nucleotide Sequen...
Nature 406, 151-157, 2000
A;Reference number: A82515; MUID:20365717; PMID:10910347
A;Note: for a complete list of authors see reference number A59328 below

RESULT 14
T35276
probable efflux protein - Streptomyces coelicolor
C;Species: Streptomyces coelicolor
C;Date: 05-Nov-1999 #sequence_revision 05-Nov-1999 #text_change 21-Jan-2000
A;Accession: T35276
R;Oliver, K.; Harris, D.; Bentley, S.D.; Parkhill, J.; Barrell, B.G.; Rajandream, M.A
submitted to the EMBL Data Library, April 1999
A;Reference number: Z21573
A;Accession: T35276
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-312 <OLI>
A;Cross-references: EMBL:AL049587; PIDN:CAB40701.1; GSPDB:GN00070; SCOEDB:SC5F2A.34c
C;Genetics:
A;Gene: SCOEDB:SC5F2A.34c
C;Superfamily: zinc transporter ZnT-2

Query Match 21.7%; Score 416.5; DB 2; Length 312;
```

Best Local Similarity 32.5%; Pred. No. 9.2e-28; Mismatches 125; Indels 29; Gaps 8;
Matches 106; Conservative 66;

Qy   52 HCHAQKGFPDSHCDPKKGKAQR-QLYVASAICLLFMIGEVVGGYLAHSLAVMTDAAHLLTD 110
         ||| ::|| :|:|||| || | | | :||::|||||||||||||||||||||:||||||
Db   7 HGHAHGG---HAGGTATAAYRGRLRVALSITILTVMVEIVGGLVADSLALIADAAHMATD 63

Qy   111 FASMLISLFSLWMSSRPATKTMNFGWQRAEILGALVSVLSIWVTGVLVYLAVERLISGD 170
         |||| : ||||| |||||||||||||| |||||||||||||||||||||||||| :| |
Db   64 ALGLGMALLAVHFASRPDSDRRIFGYARAEILAAALANCLLLGVGGVVLIYEAIDRFVTPA 123

Qy   171 YEIDGGTMLITSGCAVAVNIIMGLTLHQSGHGHSHGTINQQEENPSVRAAFIHVIGDFMQ 230
         |:|||:| ::| ||:| | ||:||||| : : :|||||||||||||||||:|::|:|::
Db   124 DYAAGPVVF--GAIGLVANMYSLSLLMRG-------------QKESLNVRGAFLEVAADAlG 171

Qy   231 SMGVLVAAYILYFKPEYKYVDPICTFVFSILVLGTTILLRDVILVLMEGTPKGVDFTAV 290
         || ||| ||| ||| ||:||| :|||||||||||||||||||||::|||: || :| |
Db   172 SLAVIVSALVI-LATGWQPADPIASLLLIGLMIVPRTLRLRETLDVLLEAAPKGVDIAEV 230

Qy   291 RDLLLSVEGVEALHSLHIWALTVAQPVLSVHIAIAQNTDAQAVLKTASSR-----LGG--- 343
         ||| |||||||||||||||||||||||||||||||||||||||||:|:|| ||
Db   231 RAHILALDGVEDVHDLHAWTISGMPVLSAHVVV----DGEALSAIGHEKMLHELQGCLG 286

Qy   344 -KFHFHVTIQIEDYSEDMKDCQACQ 368
          | : : ||:|| :: | :| |:| |
Db   287 DHFDVEHCTFQLEPSGHAEHEARLCR 312

RESULT 15
AG1396
cation transport protein (efflux) homolog lmo2575 [imported] - Listeria monocytogenes (s
C;Species: Listeria monocytogenes
C;Date: 27-Nov-2001 #sequence_revision 27-Nov-2001 #text_change 14-Dec-2001
C;Accession: AG1396
R;Glaser, P.; Frangeul, L.; Buchrieser, C.; Amend, A.; Baquero, F.; Berche, P.; Bloeck
; Dominguez-Bernal, G.; Duchaud, E.; Durand, L.; Dussurget, O.; Entian, K.D.; Fsihi, H.
D.; Jones, L.M.; Karst, U.
Science 294, 849-852, 2001
A;Authors: Kreft, J.; Kuhn, M.; Kunst, F.; Kurapkat, G.; Madueno, E.; Maitournam, A.;
ok, C.; Schlueter, T.; Simoes, N.; Tierrez, A.; Vazquez-Boland, J.A.; Voss, H.; Wehla
A;Title: Comparative genomics of Listeria species.
A;Reference number: AB1077; MUID:21537279; PMID:11679669
A;Accession: lmo2575
A;Genetics:
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-303 <GLA>
A;Cross-references: GB:NC_003210; PIDN:CAD00653.1; PID:g16412063; GSPDB:GN00177
A;Experimental source: strain EGD-e
A;Genetics:
A;Gene: lmo2575
C;Superfamily: zinc transporter ZnT-2

Query Match   21.6%; Score 414.5; DB 2; Length 303;
Best Local Similarity 30.1%; Pred. No. 1.3e-27;
Matches 93; Conservative 72; Mismatches 127; Indels 17; Gaps 5;

Qy   47 AQSNHHCHAQKGFPDSHCDPKKGKAQRQLYVASAICLLFMIGEVVGGYLAHSLAVMTDAAH 106
         : ::| : ||| :| | :| | ::
Db   2 AHNHDHAHGHNHNHAHANANKK-----SLFISFLIATFMVEVIGGIMTNSLALLSDAGH 56

Qy   107 LLTDFASMLISLFSLWMSSRPATKTMNFGWQRAEILGALVSVLSIWVTGVLVYLAVERL 166
         ||| : ||| | :| | | |||:||||||||||||||||||||||||||||||||
Db   57 MLSDAVALGLSLAAFKFGEKAASSDKTYGYKRFEILAAFLNGLTLVGISVFIFYEAIGRF 116

Qy   167 ISGDYEIDGGTMLITSGCAVAVNIIMGLTLHQSGHGHSHGTINQQEENPSVRAAFIHVIG 226
         :||| ||||:| ::| ||:| | ||:||||| | : : |||||||||||||||:|::| |
Db   117 FDPPQVIGAGMMTI-SVIGLLINILVAWILMK------GDIS---ENLNMRSAFLHVLG 165

Qy   227 DFMQSMGVLVAAYILYFKPEYKYVDPICTFVFSILVLGTTILLRDVILVLMEGTPKGVD 286
         ||:: || ||| ||| ||| ||:||| :|||||||||||||||||||||::|||: ||
Db   166 DLLGSVGAIIAALLIIFL-GWNIADPIASVIVAALILVSGWRVLKDAIHILMEGKPANVD 224

Qy   287 FTAVRDLLLSVEGVEALHSLHIWALTVAQPVLSVHIAIAQNTDAQAVLKTASSRLQGKFH 346
         | | ||| |||||||||||||||||||||||||||||||||||||||||:|:|||:| ||
Db   225 TEEIKTFFQEODGVTEVHDLHVWAITSDFNALTAHLTVAEDADRDKILTDIEHYLQENFS 284

Qy   347 FHTVTIQIE 355
         |||::|:|
Db   285 LEHSTIQLE 293

RESULT 16
AB1772
cation transport protein (efflux) homolog lin2720 [imported] - Listeria innocua (stra
C;Species: Listeria innocua
C;Date: 27-Nov-2001 #sequence_revision 27-Nov-2001 #text_change 14-Dec-2001
C;Accession: AB1772
R;Glaser, P.; Frangeul, L.; Buchrieser, C.; Amend, A.; Baquero, F.; Berche, P.; Bloec
; Dominguez-Bernal, G.; Duchaud, E.; Durand, L.; Dussurget, O.; Entian, K.D.; Fsihi,
D.; Jones, L.M.; Karst, U.
Science 294, 849-852, 2001
A;Authors: Kreft, J.; Kuhn, M.; Kunst, F.; Kurapkat, G.; Madueno, E.; Maitournam, A.;
ok, C.; Schlueter, T.; Simoes, N.; Tierrez, A.; Vazquez-Boland, J.A.; Voss, H.; Wehla
A;Title: Comparative genomics of Listeria species.
A;Reference number: AB1077; MUID:21537279; PMID:11679669
A;Accession: AB1772
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-303 <GLA>
A;Cross-references: GB:AL592022; PIDN:CAC97946.1; PID:g16415256; GSPDB:GN00178
A;Experimental source: strain Clip11262
C;Genetics:
C;Gene: lin2720
C;Superfamily: zinc transporter ZnT-2

Query Match   21.3%; Score 410.5; DB 2; Length 303;
Best Local Similarity 30.4%; Pred. No. 2.9e-27;
Matches 94; Conservative 70; Mismatches 128; Indels 17; Gaps 5;

Qy   47 AQSNHHCHAQKGFPDSHCDPKKGKAQRQLYVASAICLLFMIGEVVGGYLAHSLAVMTDAAH 106
         : ::| : ||| :| | :| | ::
Db   2 AHNHDHAHGHNHNHAHNANKK-----SLFISFLIATFMVEVIGGIMTNSLALLSDAGH 56

Qy   107 LLTDFASMLISLFSLWMSSRPATKTMNFGWQRAEILGALVSVLSIWVTGVLVYLAVERL 166
         ||| : ||| | :| | | |||:||||||||||||||||||||||||||||||||
Db   57 MLSDAVALGLSLAAFKFGEKAASSDKTYGYKRFEILAAFLNGLTLVGISVFIFYEAIGRF 116

Qy   167 ISGDYEIDGGTMLITSGCAVAVNIIMGLTLHQSGHGHSHGTINQQEENPSVRAAFIHVIG 226
         :||| ||||:| ::| ||:| | ||:||||| | : : |||||||||||||||:|::| |
Db   117 FDPPQVIGAGMMTI-SVIGLLINILVAWILMK------GDIS---ENLNMRSAFLHVLG 165

Qy   227 DFMQSMGVLVAAYILYFKPEYKYVDPICTFVFSILVLGTTILLRDVILVLMEGTPKGVD 286
         ||:: || ||| ||| ||| ||:||| :|||||||||||||||||||||::|||: ||
Db   166 DLLGSVGAIIAALLIIFL-GWNIADPIASVIVAALILVSGWRVLKDAIHILMEGKPANVD 224

Qy   287 FTAVRDLLLSVEGVEALHSLHIWALTVAQPVLSVHIAIAQNTDAQAVLKTASSRLQGKFH 346
         | | ||| |||||||||||||||||||||||||||||||||||||||||:|:|||:| ||
Db   225 TEEIKTFFQEODGVTEVHDLHVWAITSDFNALTAHLTVAEDADRDKILTDIEHYLQENFS 284

Qy   347 FHTVTIQIE 355
         |||::|:|
Db   285 LEHSTIQLE 293

RESULT 17
D89778
hypothetical protein SA0163 [imported] - Staphylococcus aureus
C;Species: Staphylococcus aureus
C;Date: 10-May-2001 #sequence_revision 10-May-2001 #text_change 22-Oct-2001
C;Accession: D89778
R;Kuroda, M.; Ohta, T.; Uchiyama, I.; Baba, T.; Yuzawa, H.; Kobayashi, I.; Cui, L.; O
ma, A.; Mizutani-Ui, Y.; Kobayashi, N.; Sawano, T.; Inoue, R.; Kaito, C.; Sekimizu, K
C.; Shiba, T.; Hattori, M.; Ogasawara, N.; Hayashi, H.; Hiramatsu, K.
Lancet 357, 1225-1240, 2001
A;Title: Whole genome sequencing of meticillin-resistant Staphylococcus aureus.
A;Reference number: A89758; MUID:21311952; PMID:11418146

A;Accession: D89778
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-319 <KUR>
A;Cross-references: GB:BA000018; PID:g13700084; PIDN:BAB41383.1; GSPDB:GN00149
A;Experimental source: strain N315
C;Genetics:
A;Gene: SA0163
C;Superfamily: zinc transporter ZnT-2

Query Match   20.8%; Score 399.5; DB 2; Length 319;
Best Local Similarity 30.1%; Pred. No. 2.6e-26;
Matches 93; Conservative 84; Mismatches 109; Indels 23; Gaps 6;

```
Qy   51  HHCHAQKGPDSHCDPKKGKAQRQLYVASAICLLFMIGEVVGGYLAHSLAVMTDAAHLLTD  110
Db   16  HHVHHQKIQQS----------SKKTLWASLIITLLFTVIEFVGGLVSNSLALLSDSFHMLSD   67

Qy  111  FASMLISFLSLWMSSRPATKTMNFGWQRAEILGALVSVLSIWWVTGVLVYLAVERLISGD  170
Db   68  VLALGLSMLAIYFASKKPTARYTFGYLRFELIVAFLNGLALIVISITWILYEAIVRIIY-P  126

Qy  171  YEIDGGTMLITSGCAVANVNIIMGLTLHQSGHGHSHGTNQQEENPSVRAAFIHVIGDFMQ  230
Db  127  QPIESGIMFMIASIGLLVNIILTIILVRS----------LKQEDNINIQSALWHFMGDLLN  177

Qy  231  SMGVLVAAYILYFKPEYKYVDPICTFVFSILVLGTTLTIILRDVILVLMEGTPKGVDFTAV  290
Db  178  SIGVIAVAVLIYF-TGWRLIDPIISIVISLIILRGGYKIFRNAWLLIMESVPQHLDTDQI  236

Qy  291  RDLLSVEGVEALHSLHIWALTVAQPVLSVHIAI---AQNTDAQAVLKTASSRLQGKFHF  347
Db  237  MADIRKNIDGILDVHEFHLWSITTEHYSLSAHVVLDKKKYEGDDYQAI-DQVSSLLKEKYGI  295

Qy  348  HTVTIQIED  356
Db  296  AHSTLQIEN  304
```

RESULT 18
C69612
cation-efflux system membrane protein czcD - Bacillus subtilis
C;Species: Bacillus subtilis
C;Date: 05-Dec-1997 #sequence_revision 05-Dec-1997 #text_change 20-Jun-2000
A;Accession: C69612
R;Kunst, F.; Ogasawara, N.; Moszer, I.; Albertini, A.M.; Alloni, G.; Azevedo, V.; Berter
C;Bron, S.; Brouillet, S.; Bruschi, C.V.; Caldwell, B.; Capuano, V.; Carter, N.M.; Cho
A;Ehrlich, S.D.; Emmerson, P.T.; Entian, K.D.; Errington, J.; Fabret, C.; Ferrari, E.
Nature 390, 249-256, 1997
A;Authors: Foulger, D.; Fritz, C.; Fujita, M.; Fujita, Y.; Fuma, S.; Galizzi, A.; Galler
iech, J.; Harwood, C.R.; Henaut, A.; Hilbert, H.; Holsappel, S.; Hosono, S.; Hullo, M.F.
Koetter, P.; Koningstein, G.; Krogh, S.; Kumano, M.; Kurita, K.; Lapidus, A.; Lardinois,
A;Authors: Lauber, J.; Lazarevic, V.; Lee, S.M.; Levine, A.; Liu, H.; Masuda, S.; Maueel
Y, M.; Ogawa, K.; Ogiwara, A.; Oudega, B.; Park, S.H.; Parro, V.; Pohl, T.M.; Portetelle
Rieger, M.; Rivolta, C.; Rocha, E.; Roche, B.; Rose, M.; Sadaie, Y.; Sato, T.; Scanlon,
A;Authors: Schleich, S.; Schroeter, R.; Scoffone, F.; Sekiguchi, J.; Sekowska, A.; Seror
akeuchi, M.; Tanaka, T.; Tanaka, K.; Terpstra, P.; Tognoni, A.; Tosato, V.; Uchiyama,
T.; Winters, P.; Wipat, A.; Yamamoto, H.; Yamane, K.; Yasumoto, K.; Yata, K.; Yoshida, K
A;Authors: Yoshikawa, H.F.; Zunstein, E.; Yoshikawa, H.; Danchin, A.
A;Title: The complete genome sequence of the Gram-positive bacterium Bacillus subtilis.
A;Reference number: A69580; MUID:98044033; PMID:9384377
A;Accession: C69612
A;Status: preliminary
A;Molecule type: nucleic acid sequence not shown; translation not shown
A;Residues: 1-311 <KUN>
A;Cross-references: GB:Z99117; GB:AL009126; NID:g2634966; PIDN:CAB14606.1; PID:g2635110
C;Genetics:
A;Gene: CzcD
C;Superfamily: zinc transporter ZnT-2

Query Match   20.7%; Score 397.5; DB 2;
Best Local Similarity 28.2%; Pred. No. 3.8e-26;
Matches 86; Conservative 86; Mismatches 118; Indels 15; Gaps 5;

```
Qy   67  KGKAQRQLYVASAICLLFMIGEVVGGYLAHSLAVMTDAAHLLTDFASMLISLFSLWMSSGR  126
Db    7  EGANKKVLLISEFIMITGTMIIEAIGGFLTNSLALLSDSISLMVALIAFTLAEK   66

Qy  127  PATKTMNFGWQRAEILGALVSVLSIWWVTGVLVYLAVERLISGDYEIDGGTMLITSGCAV  186
Db   67  KANHNKTFGYKRFEILAAVINGAALILISLVYIIYEAIER-FSNPPKVATTGMLTISIGL  125

Qy  187  AVNIIMGLTLHQSGHGHSHGTNQQEENPSVRAAFIHVIGDFMQSMGVLVAAYILYFKPE  246
Db  126  VVNLLVAWIMMSGG-----------------DTKNNLNIRGAYLHVISDMLGSVGAILIFF-G  175

Qy  247  YKYVDPICTFVFSILVLGTTLTIILRDVILVLMEGTPKGVDFPAVRDLLLSVEGVEALHSL  306
Db  176  WGNADPLASIIVAILVLRSGYNVTKDSTHILMEGTPENIDVSDIIRTIEGTEGIQNIHDL  235

Qy  307  HIWALTVAQPVLSVHIAIAQN---TDAQAVLKTASSRLQGKFHFHTVTIQIEDYSEDMKD  363
Db  236  HIWSITSGLNALSCHAVVDDQLTISESENILRKIEHELEHKGITH-VTIQMETEAHNHDN  294

Qy  364  CQACQ  368
Db  295  AILCQ  299
```

RESULT 19
AH3431
cobalt-zinc-cadmium resistance protein czcD [imported] - Brucella melitensis (strain
C;Species: Brucella melitensis
C;Date: 01-Feb-2002 #sequence_revision 01-Feb-2002 #text_change 15-Feb-2002
A;Experimental source: strain 16M
C;Genetics:
A;Accession: AH3431
A;Gene: BMEII438
A;Map position: I
R;Delvecchio, V.G.; Kapatral, V.; Redkar, R.J.; Patra, G.; Mujer, C.; Los, T.; Ivanov
.; Mazur, M.; Goltsman, E.; Selkov, E.; Elzer, P.H.; Hagius, S.; O'Callaghan, D.; Let
Proc. Natl. Acad. Sci. U.S.A. 99, 443-448, 2002
A;Title: The genome sequence of the facultative intracellular pathogen Brucella melit
A;Reference number: AD3252; PMID:11756688
A;Accession: AH3431
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-301 <KUR>
A;Cross-references: GB:AE008917; PIDN:AAL52619.1; PID:g17983439; GSPDB:GN00190
A;Experimental source: strain 16M

Query Match   20.5%; Score 393.5; DB 2; Length 301;
Best Local Similarity 30.9%; Pred. No. 7.9e-26;
Matches 93; Conservative 66; Mismatches 123; Indels 19; Gaps 6;

```
Qy   58  GPD-SHCDPKKGKAQRQLYVASAICLLFMIGEVVGGYLAHSLAVMTDAAHLLTDFASMLI  116
Db    6  GADHEHADVKNTPISR-LWIAFGLTGIFMIAEVIGSFATGSLALLSDAMHMATDAFALL   64

Qy  117  SLFSLWMSSRPATKTMNFGWQRAEILGALVSVLSIWWVTGVLVYLAVERLISGDYEIDG  176
Db   65  ALIAIYLGRRPADVLRTYGYARFEILAAAFNALLLLGVAFYLIYEAWERL-SEPADVQSL  123

Qy  177  TMLITSGCAVANVNIIMG--LTLHQSGHGHSHGTNQQEENPSVRAAFIHVIGDFMQSMGV  234
Db  124  GMLAVAIVGLVMNFISMRILTVH--------------KDDSLNVKGAYLEWADMLGSVGV  170

Qy  235  LVAAYILYFKPEYKYVDPICTFVFSILVLGTTLTIILRDVILVLMEGTPKGVDFTAVRDLL  294
Db  171  IAAAIIIYL-TGWEWVDSAIAVGIGFMVFPRIWVLLKECINILLEGVPAGGDVKKLEAAI  229

Qy  295  LSVEGVEALHSLHIWALTVAQPVLSVHIAIAQNTDAQAVLKTASSRLQGKFHFHTVTIQI  354
Db  230  LAVPGVASVHDLHVWSLIKTEHSLTAHLVLAQEADGEIVRRAVEHVLQNDYDLHHTTLQT  289
```

Qy   355 E 355
Db   290 E 290

RESULT 20
G90008
Db   G90008
A;Note: czrB [imported] - Staphylococcus aureus (strain N315)
C;Species: Staphylococcus aureus
C;Date: 10-May-2001 #sequence_revision 10-May-2001 #text_change 22-Oct-2001
C;Accession: G90008
R;Kuroda, M.; Ohta, T.; Uchiyama, I.; Baba, T.; Yuzawa, H.; Kobayashi, I.; Cui, L.; Oguma, A.; Mizutani-Ui, Y.; Kobayashi, N.; Sawano, T.; Inoue, R.; Kaito, C.; Sekimizu, K.; Shiba, T.; Hattori, M.; Ogasawara, N.; Hayashi, H.; Hiramatsu, K.
A;Title: Whole genome sequencing of meticillin-resistant Staphylococcus aureus.
Lancet 357, 1225-1240, 2001
A;Reference number: A89758; MUID:21311952; PMID:11418146
A;Accession: G90008
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-325 <KUR>
A;Gene: czrB
C;Superfamily: zinc transporter ZnT-2

Query Match   95%;   Score 391;   DB 2;   Length 325;
Best Local Similarity   28.9%;   Pred. No. 1.4e-25;
Matches   95;   Conservative   73;   Mismatches 133;   Indels   28;   Gaps   7;

```
Qy   47 AQSNHHCHAQKGPDSHCDPKKGKAQRQLYVASAICLLFMIGEVVGGYLAHSLAVMTDAAH 106
         :||:|| ||| :|:|:   |  ||||||||||||||||||||||||||:|:|
Db    2 SHSHHHDHMH----SHVTTNNNKV--LFISFLIGLYMFIEIIGGLLANSLALLSDGIH   54

Qy  107 LLTDFASMLISLFSLWMSSRPATKTMNFGWQRAEILGALVSVLSIWVVTGVLVYLAVERL 166
Db   55 MFSDTFSLGVALVAFIYAEKNATTTKFFGYKREVLAALFNGVTLFVISILIVFEAIKRF 114

Qy  167 ISGDYEIDGGTMLITSGCAVAVNIIMGLTLHQSGHGHSHGTTNQQEENPSVRAAFIHVIG 226
Db  115 FVPS-EVQSKEMLIISIIGLIVNIVVAFFMFKGG------------DTSHNLNMRGAFLHVIG 164

Qy  227 DFMQSMGVLVAAYILYFKPEYKVVDPICTFVFSILVLGTTLTLIRDVILVLMEGTPKGVD 286
Db  165 DLLGSVGAITAA-ILIWAFGWTIADPIASILVSVIILKSAWGITKSSINILMEGTPSDVD 223

Qy  287 FTAVRDLLLSVEGVEALHSLHIWALTVAQPVLSVHIAIAQNTDAQAVLKTASSRLQGKFH 346
Db  224 IDEVITTIKKDSRIQSVHDCHVWTISNDMNALSCHVVV----DHTLIMKECELLIENEH 279

Qy  347 ------FHTVTIQIEDYSEDMKDCQACQG 369
Db  280 DLLHLNIHHMTIQLETPNHKHDESIICSG 308
```

RESULT 21
T44365
Db   T44365
A;Note: cation-efflux system membrane protein homolog [imported] - Staphylococcus aureus
C;Species: Staphylococcus aureus
C;Date: 21-Jan-2000 #sequence_revision 21-Jan-2000 #text_change 11-May-2000
C;Accession: T44365
R;Kuroda, M.; Hayashi, H.; Ohta, T.
A;Title: Chromosome-determined zinc responsible operon czr in Staphylococcus aureus stra
Microbiol. Immunol. 43, 115-125, 1999
A;Reference number: Z22754; MUID:99244271; PMID:10229265
A;Accession: T44365
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-325 <KUR>
A;Cross-references: EMBL:AB016431; PIDN:BAA36686.1
C;Genetics:

A;Note: czrB
C;Superfamily: zinc transporter ZnT-2

Query Match   97%;   Score 390;   DB 2;   Length 325;
Best Local Similarity   29.3%;   Pred. No. 1.7e-25;
Matches   97;   Conservative   73;   Mismatches 129;   Indels   32;   Gaps   8;

```
Qy   47 AQSNHHCHAQKGPDSH--CDPKKGKAQRQLYVASAICLLFMIGEVVGGYLAHSLAVMTDA 104
Db    2 SHSHHHDHMH----SHVTDNKK-----VLFISFLIGLYMFIEIIGGLLANSLALLSDG   52

Qy  105 AHLLTDFASMLISLFSLWMSSRPATKTMNFGWQRAEILGALVSVLSIWVVTGVLVYLAVE 164
Db   53 IHMFSDTFSLGVALVAFIYAEKNATTTKFFGYKREVLAALFNGVTLFVISILIVFEAIK 112

Qy  165 RLISGDYEIDGGTMLITSGCAVAVNIIMGLTLHQSGHGHSHGTTNQQEENPSVRAAFIHV 224
Db  113 RFFVPS-EVQSKEMLIISIIGLIVNIVVAFFMFKGG--------DTSHNLNMRGAFLHV 162

Qy  225 IGDFMQSMGVLVAAYILYFKPEYKVVDPICTFVFSILVLGTTLTLIRDVILVLMEGTPKG 284
Db  163 IGDLLGSVGAITAA-ILIWAFGWTIADPIASILVSVIILKSAWGITKSSINILMEGTPSD 221

Qy  285 VDFTAVRDLLLSVEGVEALHSLHIWALTVAQPVLSVHIAIAQNTDAQAVLKTASSRLQGK 344
Db  222 VDIDEVITTIKKDSRIQSVHDCHVWTISNDMNALSCHVVV----DHTLIMKECELLIENI 277

Qy  345 FH------FHTVTIQIEDYSEDMKDCQACQG 369
Db  278 EHDLLHLNIHHMTIQLETPNHKHDESIICSG 308
```

RESULT 22
JC4701
Db   JC4701
A;Note: cadmium, zinc, cobalt divalent cation resistant determinant D - Alcaligenes sp.
C;Species: Alcaligenes sp.
C;Date: 10-Sep-1999 #sequence_revision 10-Sep-1999 #text_change 10-Sep-1999
C;Accession: JC4701
R;Kunito, T.; Kusano, T.; Oyaizu, H.; Senoo, K.; Kanazawa, S.; Matsumoto, S.
A;Title: Cloning and sequence analysis of czc genes in Alcaligenes sp. strain CT14.
Biosci. Biotechnol. Biochem. 60, 699-704, 1996
A;Reference number: JC4698; MUID:96219090; PMID:8829543
A;Accession: JC4701
A;Molecule type: DNA
A;Residues: 1-316 <KUN>
A;Cross-references: DDBJ:D67024
C;Genetics:
A;Gene: czcD
C;Superfamily: zinc transporter ZnT-2

Query Match   102%;   Score 387;   DB 1;   Length 316;
Best Local Similarity   34.3%;   Pred. No. 3e-25;
Matches   102;   Conservative   57;   Mismatches 120;   Indels   18;   Gaps   7;

```
Qy   61 SHCDPKKGKAQRQLYVASAICLLFMIGEVVGGYLAHSLAVMTDAAHLLTDFASMLISLFS 120
Db    6 SHDHP--GGNERSIKIALALTGTFILAEVVGGVMTKSLALISDAAHMLTDTVALAIALAA  63

Qy  121 LWMSSRPATKTMNFGWQRAEILGALVSVLSIWVVTGVLVYLAVERLISGDYEIDGGTMLI 180
Db   64 IAIAKRPADKKRTFGYKRFEILAAAFNALLLFGVAIYILYEAYLRLKSPP-QIESTGMFV 122

Qy  181 TSGCAVAVNIIMGLTLHQSGHGHSHGTTNQQEENPSVRAAFIHVIGDFMQSMGVLVAAYI 240
Db  123 VAVLGIIINLI-SMRMLSSG---------QSSSLNVKGAYLEWSDLLGSVGVIAGAII 171

Qy  241 LYFKPEYKVVDPICTFVFSILVLGTTLTLIRDVILVLMEGTPKGVDFTAVRDLLLSVEGV 300
Db  172 IRF-TGWAWVDSAIAVLIGLWVLPRTWFLKSSLNVLEGVPDDVDLAEVEKQILATPGV 230

Qy  301 EALHSLHIWALTVAQPVLSVHIA---IAQNTDAQAVLKTASSRLQGKFHFHHTVTIQIE 355
Db  231 KSFHDLHIWALTSGKASLIVHVNDTAVNPEME-VLPELKQMLADKFDITHVTIQFE 286
```

RESULT 23

AI0138

probable cation transport protein YP01129 [imported] - Yersinia pestis (strain CO92)
C;Species: Yersinia pestis
C;Date: 02-Nov-2001 #sequence_revision 02-Nov-2001 #text_change 09-Nov-2001
C;Accession: AI0138
R;Parkhill, J.; Wren, B.W.; Thomson, N.R.; Titball, R.W.; Holden, M.T.G.; Prentice, M.B.;
deno-Tarraga, A.M.; Chillingworth, T.; Cronin, A.; Davies, R.M.; Davis, P.; Dougan, G.;
11, M.; Rutherford, K.; Simmonds, M.; Skelton, J.; Stevens, K.; Whitehead, S.; Barrell,
Nature 413, 523-527, 2001
A;Title: Genome sequence of Yersinia pestis, the causative agent of plague.
A;Reference number: AI0138
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-312 <KUR>
A;Cross-references: GB;AL590842; PIDN:CAC89972.1; PID:g15979196; GSPDB:GN00175
C;Genetics:
A;Gene: YP01129
C;Superfamily: zinc transporter ZnT-2

Query Match   19.8%;   Score 380.5;   DB 2;   Length 312;
Best Local Similarity   30.4%;   Pred. No. 1.1e-24;
Matches   92;   Conservative   71;   Mismatches   125;   Indels   15;   Gaps   6;

```
Qy     72 RQLYVASAICLLFMIGEVVGGYIAHSLAVMTDAAHLLTDFASMLISLFSLWMSSRPATKT 131
             :|| || | |||:| ::||  |:|| |:||||||||||| :|:| |:|:| ||:  || |
Db     14 KRLIIAFAITTLFMVTEAIGGWLSGSLLALLADAGHMLTDSAALFIALMAVHFSQRKPDPR 73

Qy    132 MNFGWQRAEILGALVSVLSIWVTVGVLVLYLAVERLISGDYEIDGGTMLITSGGCAVANII 191
             | ||:|||||| |||:|| || ||::||| |::| :|| ||| :|| |||:|| | || |
Db     74 HTFGYLRLTLAAFVNAAALLIVILIVWEAVHRFFS-PHEVWGTPMLIIAIAGLLANIF 132

Qy    192 MGLTLHQSGHGHSHGTTNQQEENPSVRAAFIHVIGDEMQSMGVLVAAYILYFKPEYKYVD 251
             |||||| | || || |||||||||||||||||||||||||||||||||||||| | | |
Db    133 CFWILHK--------GEEEKNINVRAAALHVLSDLLGSVGAMIAA-IVILLTTGWTPID 181

Qy    252 PICTFVFSILVLGTTLTLILRDVILVLMEGTPKGVDFTAVR-DLLSVEGVEALHSLHIWA 310
             |||| | :||| |||||:|||||||||||||||||||||||  ||| ||| | ||||||
Db    182 PILSLVLVSVLLRSAWRLLKESSFHELLEGAPOEIDINKLRKDLCTNIYEVRNIHHVHLWQ 241

Qy    311 LTVAQPVLSVHIAIAQNTDAQAVLKTASSRLQGKFHFHTVTIQIEDYSEDMKDCQACQ-G 369
             | || | :::|   ||| |||:|||||||:| :|| |:| |:||||||||||||||| |
Db    242 VG-EQRLMTLHAQVIPPLDHDALLQRIQDYLLHHYRISHATVQMEYQHCGTPDCGINQAA 300

Qy    370 PSD 372
             | |
Db    301 PAD 303
```

RESULT 24

F87286

cation efflux family protein [imported] - Caulobacter crescentus
C;Species: Caulobacter crescentus
C;Date: 20-Apr-2001 #sequence_revision 20-Apr-2001 #text_change 20-Apr-2001
C;Accession: F87286
R;Nierman, W.C.; Feldblyum, T.V.; Paulsen, I.T.; Nelson, K.E.; Eisen, J.; Heidelberg, J.
B.; Laub, M.T.; DeBoy, R.T.; Dodson, R.J.; Durkin, A.S.; Gwinn, M.L.; Haft, D.H.; Kolon
n, J.; Ermolaeva, M.; White, O.; Salzberg, S.L.; Shapiro, L.; Venter, J.C.; Fraser, C.M.
Proc. Natl. Acad. Sci. U.S.A. 98, 4136-4141, 2001
A;Title: Complete Genome Sequence of Caulobacter crescentus.
A;Reference number: A87249; MUID:21173698; PMID:11259647
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-361 <STO>
A;Cross-references: GB;AE005673; NID:g13421446; PIDN:AAK22290.1; GSPDB:GN00148
C;Genetics:
A;Gene: CC0303

Query Match   19.5%;   Score 374.5;   DB 2;   Length 361;

Best Local Similarity   28.3%;   Pred. No. 4e-24;
Matches   92;   Conservative   70;   Mismatches   140;   Indels   23;   Gaps   8;

```
Qy     50 NHHCHAQKGPDSHCDPKKGGKAQRQLYVASAICLLFMIGEVVGGYIAHSLAVMTDAAHLLT 109
             | ||  |  :||  ||| | |:||| |:||::|||:||| ||||| ||| |:|| |||||
Db     56 HHHGHGHGHHHHHAPKDFGRA--FAIGTALNLGFVIVEATAGILTHSLALLADAGHNLS 112

Qy    110 DFASMLISLFSLWMSSRPATKTMNFGWQRAEILGALVSVLSIWVTVGVLVLYLAVERLISG 169
             ||||||:||||||||| |||||  ||:|||||| |||::|| || ||::||| |::| :||
Db    113 DVLGLLANGATVLAKRAPSARRIYGLRKGTIASLGNAALLVAVGAIAWEGVRRFAAP 172

Qy    170 DYEIDGGTMLITSGCAVANIIMGLTLHQSGHGHSHGTTNQQEENPSVRAAFIHVIGDPM 229
             || :|| |||:|| | || || |||||||| || |||||||||||||||||||||||||
Db    173 E-PVQTGPMIVAAIGIVINTATALMFMKG---------SKEDLNVRGAFLHMAADAA 220

Qy    230 QSMGVLVAAYILYFKPEYKYVDPICTFVP-SILVLGTTLTLILRDVILVLMEGTPKGVDFT 288
             ||||||||||||||||| | | ||| | |  |:||| |||||||||||||||||||||||
Db    221 VSAGVVIAALAMTF-TGWNWLDPVVSLVIVAVIVLG-TWGLLRDSLIDLALDATPRGIDTQ 278

Qy    289 AVRDLLSVEGVEALHSLHIWALTIVAQPVLSVHIAIAQNTDAQAVLKTASSRLQGKFHFH 348
             |||| |||||||| ||||||| |||||| | ||:||| ||| |||::||||:| :|| |
Db    279 KVRDWLAARPGVSEVHDLHIWAMSTTETALTAHVVRQLDADHDQFLHDACAELASRFNIG 338

Qy    349 TVTIQIEDYSEDMKDCQACQ-GPSD 372
             |:| |:|||||||||||||| | |
Db    339 HVTIQV---ESGHGAHACRLAPAD 359
```

RESULT 25

C83595

probable cation efflux system protein PA0397 [imported] - Pseudomonas aeruginosa
C;Species: Pseudomonas aeruginosa (str
C;Date: 15-Sep-2000 #sequence_revision 15-Sep-2000 #text_change 31-Dec-2000
C;Accession: C83595
R;Stover, C.K.; Pham, X.Q.; Erwin, A.L.; Mizoguchi, S.D.; Warrener, P.; Hickey, M.J.;
adman, S.; Yuan, Y.; Brody, L.L.; Coulter, S.N.; Folger, K.R.; Kas, A.; Larbig, K.; L
.; Lory, S.; Olson, M.V.
Nature 406, 959-964, 2000
A;Title: Complete genome sequence of Pseudomonas aeruginosa PA01, an opportunistic pa
A;Reference number: A82950; MUID:20437337; PMID:10984043
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-299 <STO>
A;Cross-references: GB;AE004477; GB:AE004091; NID:g9946248; PIDN:AAG03786.1; GSPDB:GN
A;Experimental source: strain PA01
C;Genetics:
A;Gene: PA0397
C;Superfamily: zinc transporter ZnT-2

Query Match   19.3%;   Score 372;   DB 2;   Length 299;
Best Local Similarity   32.2%;   Pred. No. 5.3e-24;
Matches   88;   Conservative   62;   Mismatches   109;   Indels   14;   Gaps   5;

```
Qy     84 FMIGEVVGGYIAHSLAVMTDAAHLLTDFASMLISLFSLWMSSRPATKTMNFGWQRAEILG 143
             |:|||||||||||||||||||||||||||||||||||||||||||||  ||:|||||||
Db     27 FLVAEVVGGIILTGSLALISDAAHMLTDAVALALAAINIARRPTNDRLITYGVHRFELLA 86

Qy    144 ALVSVLSIWVTVGVLVLYLAVERLISGDYEIDGGTMLITSGCAVANIIMGLTLHQSGHGH 203
             |||::|| || || ::||| |::| :||  :|| |||:|| | || || |||||||| ||
Db     87 AAFNAFLFGVAFYILYAAYERL-NQPAEIQSVGMLVIAVLGLLVN-LASMRLLAPAQGN 144

Qy    204 SHGTTNQQEENPSVRAAFIHVIGDFMQSMGVLVAAYILYFKPEYKYVDPICTFVFSILVL 263
             || |:| ::  ||||||||||||| ||||||||||||||||||| | | |||| | :|||
Db    145 S---------LNVKGAYLEVWSDMLGSLGVIVAAIVIRF-TGWAWVDSLVAVLIGFWVL 193

Qy    264 GTTLTLILRDVILVLMEGTPKGVDFTAVRDLLSVEGVEALHSLHIWALTIVAQPVLSVHIA 323
             |||||:|||||||||||||||||||:||||| ||| |||| |||||| |||||| | |||
Db    194 PRIWILRESLHVLLEGVPKEIQLAELREALGIPGVTGLHDLHVWSITSGKISLTSHLV 253

Qy    324 I-AQNTDAQAVLKTASSRLQGKFHFHTVTIQIE 355
             : || | |:|||:||||:| :||||||:| |:|
Db    254 YDPALVDAEALLGTVKALLHDRYEIEHSTLQLE 286
```

RESULT 26
S56057
heavy metal ion resistance protein ZRC1 - yeast (Saccharomyces cerevisiae)
N;Alternate names: protein YM9408.05c; protein YMR243c
C;Species: Saccharomyces cerevisiae
C;Date: 27-Aug-1995 #sequence_revision 01-Mar-1996 #text_change 21-Jul-2000
C;Accession: S56057; JQ0349
R;Gentles, S.; Bowman, S.
submitted to the EMBL Data Library, March 1995
A;Reference number: S56057
A;Accession: S56057
A;Molecule type: DNA
A;Residues: 1-442 <GEN>
A;Cross-references: EMBL:Z48756; NID:g736304; PIDN:CAA88653.1; PID:g736309; MIPS:YMR243c
R;Kamizono, A.; Nishizawa, M.; Teranishi, Y.; Murata, K.; Kimura, A.
Mol. Gen. Genet. 219, 161-167, 1989
A;Title: Identification of a gene conferring resistance to zinc and cadmium ions in the
A;Reference number: JQ0349; MUID:90136503; PMID:2693940
A;Accession: JQ0349
A;Molecule type: DNA
A;Residues: 1-413,'S',415-416,'I',418-442 <KAM>
A;Cross-references: EMBL:X17537; NID:g4835; PIDN:CAB56542.1; PID:g5924031
C;Genetics:
A;Gene: SGD:ZRC1
A;Cross-references: SGD:S0004856; MIPS:YMR243c
A;Map position: 13R
C;Function:
A;Description: cadmium resistance; zinc resistance
C;Keywords: transmembrane protein
F;9-25/Domain: transmembrane #status predicted <TM1>
F;42-58/Domain: transmembrane #status predicted <TM2>
F;80-96/Domain: transmembrane #status predicted <TM3>
F;113-129/Domain: transmembrane #status predicted <TM4>
F;240-256/Domain: transmembrane #status predicted <TM5>
F;270-286/Domain: transmembrane #status predicted <TM6>

Query Match   18.6%;   Score 357; DB 2;   Length 442;
Best Local Similarity   25.2%;   Pred. No. 1.6e-22;
Matches   97;   Conservative 80;   Mismatches 108;   Indels 100;   Gaps 13;

Qy   72 RQLYVASAICL--LFMIGEVGVGGYLAHSLAVMTDAAHLLTDFASMLISLFSLWMS-SRPA 128
               :||||::|| ||:::|| :|||||:::||||:||:||:|:|:||:||:: :||
Db   5 KELRIISLLTLDTVFFLLEITIGYMSHSLALIADSFHMLNDIISLLVALWAVDVAKNRGP 64

Qy   129 TKTMNFGWQRAEILGALLVSVLSIWVVTGVLVLVLAVERLISGDYEIDGGTMLITSGCAVAV 188
               ||:||||||||||:|:: :| :||||||:||:|||||:||||:|| :
Db   65 DAKYTYGWKRAEILGALINAVFLIALCFSIMIEALQRLIE-PQEIQNPRLVLVGVAGLI 123

Qy   189 NIIMGLTL-HQSG----HGHSGTTN------------------------------- 209
               :||||:|| ||: ||||||||
Db   124 SNVVGLFLFHDHGDSDLHSHSHGSVESGNNDLDIESNATHSHSHASLPNDNLAIDEDAIS 183

Qy   210 -----------------QQEENP--------------------SVRAAFIHVIGDF 228
                                | ||:|| :||| :||:|||
Db   184 SPGPSGQIGEVLPQSVVNRLSNESQPLLNHDHDHSHESKKPGHRSLNMHGVFLHVLGDA 243

Qy   229 MQSMGVLVAAYILYFKPEYK----YVDPICTFVFSILVLGTTLTILRDVILVLMEGTPKGV 285
               |||||||:|||:||||||| :| ||| |||:||||||:|:||||||:|||||||||
Db   244 LGNIGVIAAALFIW-KTEYSWRYSDPIVSLIITIIIFSSALPLSRRASRILLQATPSTI 302

Qy   286 DFTAVRDLLSVEGVEALHSLHIWALTVAQPVLSVHIAIAQNTDAQAVLKTASSRLQGK- 344
               ||||:|||||||||||||||||||||||||||||||| | :|||:|: :|:||:||||
Db   303 SADQIQREILAVPGVIAVHDFHVWNLTESIYIASIHVQI----DCAPDKFMSSAKLIRKI 358

Qy   345 FH---FHHVTIQIE----DYSEDMK 362
               |:| |::|:|:| |:||||
Db   359 FHQHGIHSATVQPEFVSGDVNEDIR 383

RESULT 27
A75437

Query Match   18.0%;   Score 346.5; DB 2;   Length 308;
Best Local Similarity   27.2%;   Pred. No. 8e-22;

---

cation efflux system protein - Deinococcus radiodurans (strain R1)
C;Species: Deinococcus radiodurans
C;Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 17-Mar-2000
C;Accession: A75437
R;White, O.; Eisen, J.A.; Heidelberg, J.F.; Hickey, E.K.; Peterson, J.D.; Dodson, R.J
, M.; Shen, M.; Vamathevan, J.J.; Lam, P.; McDonald, L.; Utterback, T.; Zalewski, C.;
, M.; Smith, H.O.; Venter, J.C.; Fraser, C.M.
Science 286, 1571-1577, 1999
A;Title: Genome sequence of the radioresistant bacterium Deinococcus radiodurans R1.
A;Reference number: A75250; MUID:20036896; PMID:10567266
A;Accession: A75437
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-325 <WHI>
A;Cross-references: GB:AE001960; GB:AE000513; NID:g6458833; PIDN:AAF10676.1; PID:g645
A;Experimental source: strain R1
C;Genetics:
A;Gene: DR1102
A;Map position: 1
C;Superfamily: zinc transporter znr-2

Query Match   18.3%;   Score 351; DB 2;   Length 325;
Best Local Similarity   28.1%;   Pred. No. 3.5e-22;
Matches   87;   Conservative 73;   Mismatches 130;   Indels 20;   Gaps 5;

Qy   48 QSNHHCHAQKGPDSHC--DPKKGKRAQRQLYVASAICLLFMIGEVGVGGYLAHSLAVMTDAA 105
               | :||| : | ||| | ||||:||:::||||:||||||| :||||:|||::
Db   21 QEHGHAHGHAHGHDHAAHNHAAGAGEROLTGALVLTGAFLVLEVAYALSSRSLALLSDAG 80

Qy   106 HLLTDFASMLISLFSLWMSSRPATKTMNFGWQRAEILGALLVSVLSIWVVTGVLVLVLAVER 165
               ||||||||||:||| : :::||| |||| |||||||||:|:|| :| :|||:|: ||
Db   81 HMLTDVAALALALFAIRMGRRPADRQRTFGYRRTFEVLAAALNAGALFAIGLYILMEAVQR 140

Qy   166 LISGDYEIDGGTMLITSGCAVAVNIIMGLTLHQSGHGHSHGTTNQQEENPSVRAAFIHVI 225
               ||||::| : :|| |||||::| || ||:|||| ::| | :|||:|||:|
Db   141 -FRQPVEVQQTTSMLVVAVAGLVVNLLSARLLAGG--------EGLNLRAAYLEVL 186

Qy   226 GDFMQSMGVLVAAYILYFKPEYKYVDPICTFVFSIIVLGTTLTILRDVILVLMEGTPKGV 285
               ||| ||||| |||||||||||:|||| ||||||::|:||||||:|:||||||||||||||
Db   187 GDLLGSVAVIAGALLIRL-TGWSWVDPLLGAGIGLWVLPRTWSLLKTSVNVLLEGVPEGL 245

Qy   286 DFTAVRDLLSVEGVEALHSLHIWALTVAQPVLSVHIAIAQNTDAQAVLKTASSRLQGKF 345
               ||||:|||||||||||||||||||:|:::|| | ||| | :|| |: :|:|| :|
Db   246 DLDALRAELRALPGVQDVHDLHVWSVTGGVVNLTAHLV---SDRAPAELLPAVHEVAHGA 302

Qy   346 HFHVTIQIE 355
               || |:|:|
Db   303 GIEHVTQVE 312

RESULT 28
E70392
cation efflux system (czcD-like) - Aquifex aeolicus
C;Species: Aquifex aeolicus
C;Date: 08-May-1998 #sequence_revision 08-May-1998 #text_change 20-Sep-1999
C;Accession: E70392
R;Deckert, G.; Warren, P.V.; Gaasterland, T.; Young, W.G.; Lenox, A.L.; Graham, D.E.;
V.
Nature 392, 353-358, 1998
A;Title: The complete genome of the hyperthermophilic bacterium Aquifex aeolicus.
A;Reference number: A70300; MUID:98196666; PMID:9537320
A;Accession: E70392
A;Status: preliminary; nucleic acid sequence not shown; translation not shown
A;Molecule type: DNA
A;Residues: 1-308 <AQF>
A;Cross-references: GB:AE000721; NID:g2983544; PIDN:AAC07126.1; PID:g2983552; GB:AE00
A;Experimental source: strain VF5
C;Genetics:
A;Gene: czcD
C;Superfamily: zinc transporter znr-2

Matches   82;   Conservative   80;   Mismatches   119;   Indels   21;   Gaps   6;

```
Qy    66  KKGKAQRQLYVASAICLLFMIGEVVGGYLAHSLAVWMTDAAHLLTDFASMLISLFSLWMSS 125
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     2  EREKSLKVLAFSFLLIFLFAFIEFLGGLLTNSLALLSDAGHMLTDAVSLSISTALVAQYLAL 61

Qy   126  RPATKTMNFGWQRAEILGALVSVLSIWWVTGVLVIYLAVERLISGDYEIDGGTMLITSGCCA 185
          ||| | | |||||||||||||||||| | || | | |||||||| || |||| |
Db    62  KVKTKRTYGLYRLEVLAALVNGVFLLGLIGYIILEAIHRF--ENEPVKPQMIYIYAFAG 119

Qy   186  VAVNIIMGLTILHQSGHGHSHGSHGTTNQQEENPSVRAAFIHVIGDFMQSMGVLVAAAYILYFKP 245
          | : | | | |||||||||||| || | | || ||||| ||| | ||
Db   120  LIVNLVVGYILLK----HS----EENINIKSALLHVAVTDLGSVAAIIAGIAIVFWK 168

Qy   246  EYKYRVDPICTFVFSILVLGTTLTILRDVILVLMEGTPKGVDFTAVRDLLLSVEGVEAHLHS 305
          | | ||||| | : : ||| | | |||| || | | | || | || || | |
Db   169  FY-LADPILSVAVALILLPSAYSVIKETVNVLLEVAPSHINTEELEKELLNLQGVKGVHD 227

Qy   306  LHIWALTVAQPVLSVHIAIAQNTDAQAVLKTASSRLQGKFHFHTVTIQIEDYSEDMKDCQ 365
          ||| | | |||| | |: | |||||||||||||||| | | |
Db   228  LHVWSITPGIEVLTVHVVVEDTSICNDILKEV-EKIAHKYGIKHTTVQLE----KEGY 280

Qy   366  AC 367
          ||
Db   281  AC 282
```

RESULT 29

S58327

cobalt accumulation protein COT1 - yeast (Saccharomyces cerevisiae)

N;Alternate names: protein O61J1; protein YOR316c
A;Species: Saccharomyces cerevisiae
C;Date: 13-Jan-1996 #sequence_revision 01-Mar-1996 #text_change 29-Oct-1999
A;Accession: S58327; S31302; S67222; S71996
R;Pearson, B.M.; Hernando, Y.; Wolf, S.S.; Kalogeropoulos, A.; Schweizer, M.
submitted to the EMBL Data Library, August 1995
A;Reference number: S58318
A;Accession: S58327
A;Molecule type: DNA
A;Residues: 1-439 <PEA>
A;Cross-references: EMBL:X90565; NID:g940836; PIDN:CAA62171.1; PID:g940847
R;Conklin, D.S.; McMaster, J.A.; Culbertson, M.R.; Kung, C.
Mol. Cell. Biol. 12, 3678-3688, 1992
A;Title: COT1, a gene involved in cobalt accumulation in Saccharomyces cerevisiae.
A;Reference number: S31302
A;Accession: S31302; MUID:93075034; PMID:1508175
A;Molecule type: DNA
A;Residues: 1-226, 'E', 228-332, 'RV', 335-423, 'V', 425-439 <CON>
A;Cross-references: EMBL:M88252; NID:g171262; PIDN:AAA74884.1; PID:g171263
R;Pearson, B.M.; Hernando, Y.; Kalogeropoulos, A.; Schweizer, M.
submitted to the Protein Sequence Database, July 1996
A;Reference number: S67213
A;Accession: S67222
A;Molecule type: DNA
A;Residues: 1-439 <PEW>
A;Cross-references: EMBL:Z75224; NID:g1420693; PIDN:CAA99636.1; PID:e252147; PID:g1420691
A;Experimental source: strain S288C
R;Pearson, B.M.; Hernando, Y.; Payne, J.; Wolf, S.S.; Kalogeropoulos, A.; Schweizer, M.
Yeast 12, 1021-1031, 1996
A;Title: Sequencing of a 35.71 kb DNA segment on the right arm of yeast chromosome XV re...
A;Reference number: S71986; MUID:97051589; PMID:8896266
A;Accession: S71996
A;Molecule type: DNA
A;Status: nucleic acid sequence not shown; translation not shown
A;Molecule type: DNA
A;Residues: 1-439 <PEF>
A;Cross-references: EMBL:X90565; NID:g940836; PIDN:CAA62171.1; PID:g940847
A;Note: the nucleotide sequence was submitted to the EMBL Data Library, August 1995

A;Description: appears to act in a dosage-dependent manner to counteract the adverse
conditions and may be important in the supply of metal that is required for metalloen
C;Keywords: mitochondrial inner membrane; mitochondrion; transmembrane protein
F;10-27/Domain: transmembrane #status predicted <TM1>
F;43-60/Domain: transmembrane #status predicted <TM2>
F;78-100/Domain: transmembrane #status predicted <TM3>
F;114-133/Domain: transmembrane #status predicted <TM4>
F;244-265/Domain: transmembrane #status predicted <TM5>
F;279-295/Domain: transmembrane #status predicted <TM6>

Query Match   17.5%;   Score 337;   DB 2;   Length 439;
Best Local Similarity   23.8%;   Pred. No. 7.7e-21;
Matches   91;   Conservative   83;   Mismatches   104;   Indels   11;   Gaps   11;

```
Qy    72  RQLYVASAICL--LFMIGEVVGGYLAHSLAVWMTDAAHLLTDFASMLISLFSLWMSS-SRPA 128
          | | |: |: | ||||||||||| ||||| | | |
Db     6  KQVKIISLLLLDIVFFGIEITTGYLSHSLALIADSFHMLNDIISLVVALWAVNAVAKNRNP 65

Qy   129  TKTMNFGWQRAEILGALVSVLSIWWVTGVLVIYLAVERLISGDYEIDGGTMLITSGCCAVV 188
          :||||||||||||||||||||||| | :| ||||||||||||||| | |
Db    66  DSTYTYGWKRAEILGALINAVFLIALCVSILEALQRIIAPPV-IENPKFVILYGVAGLI 124

Qy   189  NIIMGLTL-----HQSGHGHSHG---------------TT----- 208
          : | | |||| ||||| | | ||
Db   125  SNTVGLFLFHDNDQEHGHGHSHGGIFADHEMMMPSSHTHTHAHVDGIENTTPMDSTDN 184

Qy   209  --------------------------------------------NQQEENPSVRAA 220
          : ||| |||| |
Db   185  ISEIMPNAIVDSFMNENTRLLTPENASKTPSYSYSSSHHTIASGGNYTEHNKRKSLNMIGV 244

Qy   221  FIHVIGDFMQSMGVLVAAYILYFKPEYK---YDPICTFVFSILVLGTTLTILRDVILVL 277
          |: |||:|| || || | | | ||| | || || :|
Db   245  FLHVLGDALGNIGVMLSAFFIW-KTDYSWKYYTDPLVSLIITGIIFSSALPLSCKASKIL 303

Qy   278  MEGTPKGVDFTAVRDLLLSVEGVEALHSLHIWALTVAQPVLSVHIAI---AQNTDAQAV 333
          ||| ||| | || | || | :|| ||| ||| | | | |: :|
Db   304  LQATPSTLSGDDVEGDLLKIPGIIAIHDFHIWNLTESIFIASLHIQLDISPEQFTD---L 360

Qy   334  LKTASSRLQGKFHFHTVTIQIE 355
          : || |||||||||||||| |
Db   361  AKIVRSKLH-RYGHSATLQPE 381
```

RESULT 30

H97212

co/Zn/Cd efflux system component [imported] - Clostridium acetobutylicum

A;Species: Clostridium acetobutylicum
C;Date: 14-Sep-2001 #sequence_revision 14-Sep-2001 #text_change 30-Sep-2001
A;Accession: H97212
R;Nolling, J.; Breton, G.; Omelchenko, M.V.; Markarova, K.S.; Zeng, Q.; Gibson, R.; L
.; Daly, M.J.; Bennett, G.N.; Koonin, E.V.; Smith, D.R.
J. Bacteriol. 183, 4823-4838, 2001
A;Title: Genome Sequence and Comparative Analysis of the Solvent-Producing Bacterium
A;Reference number: H97212
A;Accession: A96900; MUID:21359325; PMID:21359325
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-303 <KUR>
A;Cross-references: GB:AE001437; PIDN:AAK80491.1; PID:g15025562; GSPDB:GN00168
A;Experimental source: Clostridium acetobutylicum ATCC824
C;Genetics:
A;Gene: CAC2540
C;Superfamily: zinc transporter ZnT-2

Query Match   17.5%;   Score 336.5;   DB 2;   Length 303;
Best Local Similarity   25.0%;   Pred. No. 5.6e-21;
Matches   82;   Conservative   86;   Mismatches   129;   Indels   31;   Gaps   7;

```
Qy    43  IELAAQSNHHCHAQKGPDSHCDPKKGKAQRQLYVASAICLLFMIGEVVGGYLAHSLAVWT 102
          :| || : |||: ||| || | ||||||||||||||
Db     1  MEKHSHEHHHH-------------DISSISKNLFFVIIFNFIIALAEVIGGSIGSGLSLIS 50

Qy   103  DAAHLLTDFASMLISLFSLWMSSRPATKTMNFGWQRAEILGALVSVLSIWWVTGVLVLYLA 162
```

RESULT 31
S54303
zinc transport protein ZnT-1 - rat
N;Alternate names: zinc transporter ZnT-1
C;Species: Rattus norvegicus (Norway rat)
C;Date: 15-Jul-1995 #sequence_revision 01-Sep-1995 #text_change 05-Nov-1999
C;Accession: S54303
R;Palmiter, R.D.; Findley, S.D.
EMBO J. 14, 639-649, 1995
A;Title: Cloning and functional characterization of a mammalian zinc transporter that co
A;Reference number: S54302; MUID:95188868; PMID:7882967
A;Accession: S54303
A;Status: preliminary
A;Molecule type: mRNA
A;Residues: 1-507 <PAL>
A;Cross-references: EMBL:U17133; NID:g577842; PIDN:AAA79234.1; PID:g577843

Query Match   17.4%;   Score 335; DB 2; Length 507;
Best Local Similarity   24.0%;   Pred. No. 1.3e-20;
Matches 105; Conservative 73; Mismatches 118; Indels 142; Gaps 13;

```
Qy    51 DALHNLSDTASLILSYVSIKISEKPKNKAKTYGYKRANILAAFINSAAL---IGISIFLG 107
          || ||| |||:|  : |:||:||:||:|:|   |  | ||:||:||:|      |:  |:|
Db   275 EHLINEDPHITHVVIQFSTSCEDNICK 302
```

```
Qy   163 VERL--ISGDYEIDGGTMLITSGCAVAVNIIMGLTLHOSGHGHSHGTTNQQEENPSYRAA 220
          |:||   ||| |::|  ::||:|:||||:||: | :||:|||  |:::||::
Db   108 VQALEKLITSLKINANIVIVVALIGLLGNFLSVIILKKGA-----EKSLNVRSS 156
```

```
Qy   221 FIHVIGDFMQSMGVLVA--AYILLYFKPEYKYVDPICTFVFSILVLGTTLTILRDVILVLME 279
          ||:||||||:|||||||  |:|||:|  |||||||:|:|:||||||:||||:||||:|||:|
Db   157 YLHMLSDAMASLAVIISGVFIKYF--AIYWIDSVLTIFINMLILKSSYNILKESINILMQ 214
```

```
Qy   280 GTPKGVDFTAVRDLLLSVEGVEALHSLHIWALTVAQPVLSVHIAIAQ--NTDAQAVLKTA 337
          |||:|||||||||||:||||||||||||||||||:|||||:|:  :|   || |||:|:|
Db   215 TTPVNLDMDDVKEQLLKEIKEIKGVHHFHIWTLDENNIVLEGHIEIDDILVSETRAISDKI 274
```

```
Qy   338 SSRLQGKFHFHVTIQIEDYSEDMKDCQ 365
          |  | |:|||||:||||:|||||||:|
Db   275 EHLINEDPHITHVVIQFSTSCEDNICK 302
```

RESULT 32
G72363
cation efflux system protein - Thermotoga maritima (strain MSB8)
C;Species: Thermotoga maritima
C;Date: 11-Jun-1999 #sequence_revision 11-Jun-1999 #text_change 21-Jul-2000
C;Accession: G72363
R;Nelson, K.E.; Clayton, R.A.; Gill, S.R.; Gwinn, M.L.; Dodson, R.J.; Haft, D.H.; Hic
Garrett, M.M.; Stewart, A.M.; Cotton, M.D.; Pratt, M.S.; Phillips, C.A.; Richardson,
C.M.
Nature 399, 323-329, 1999
A;Title: Evidence for lateral gene transfer between Archaea and Bacteria from genome
A;Reference number: A72200; MUID:99287316; PMID:10360571
A;Accession: G72363
A;Molecule type: DNA
A;Residues: 1-284 <ARN>
A;Cross-references: GB:AE001729; GB:AE001512; NID:g4981049; PIDN:AAD35623.1; PID:g498
A;Experimental source: strain MSB8
A;Status: preliminary
A;Gene: TM0538
C;Genetics:
A;Genetics:
C;Superfamily: zinc transporter

Query Match   17.4%;   Score 334.5; DB 2; Length 284;
Best Local Similarity   28.2%;   Pred. No. 7.6e-21;
Matches 81; Conservative 73; Mismatches 108; Indels 25; Gaps 8;

```
Qy    68 GKAQRQLVASAICLLFMIGEVVGGYLAHSLAVMTDAAHLLTDFASMLISLFSLWMSSRP 127
          |: |:|: : |  ||  :|| ||  |||:|   |||:||:|||::| ||:|  :| |
Db     5 GRNRGRLLCMLLITFMFNVLEVVVSRVTASLAMLSDSFHMLSDVIALVVALVAERFARRT 64
```

```
Qy   128 -AIKTMNFGWQRAEILGALVSLVSIWWTGVLVYLAVERLISGDYEIDGGTMLITSGCAV 186
           |: |:| |: :|  |  || : |: | : || :| |:::|:|  :| :||::|:||:
Db    65 HAYQKNTPFGWIRAEVMGALVNALFLTGLCFAILLEAVERFIE-PHEMQQPLVLSVGVAG 123
```

```
Qy   187 AVNIIMGLTL--HQS------GHGHSHG----------- 206
          |:|:|:||||  |:|      |||| |
Db   124 LLVNVLGCLFHHHSGEGQGACHGHSHGHGHLAKGARKAGRAGGEAGAPPGRAPDQEP 183
```

```
Qy   207 --------------------------TTN------------------QQEENP----S 216
                                      :||                     |:| |     :
Db   184 DQEETNILVANISNSGLKADQAEPEKLRSDDPVDVQVNGNLIQEBSDSLESEDNRAGQLN 243
```

```
Qy   217 VRAAFIHVIGDFMQSMGVLVAAYILYFK-------------------P 245
          | ||||:||||||:|||||||:|||||:                   :
Db   244 MRGVFLHVLGDALGSVIVVNALVFYFSWKGCTEDDFCVNPCFPDPCKSSVELMNSTQAP 303
```

```
Qy   246 EYK--------YVDPICTFVFSILVLGTTLTILRDVILVLMEGTPKGVDFTAVRDLLLSVE 298
          |:|        :| |:| ||||||:||||:|||||||::|::| |:|||||:||||||:|
Db   304 MHEAGPCWVLYLDPTLCIIMVCILLYTTYPLLKESALILLQTVPKQIDIKHLVKELRDVE 363
```

```
Qy   299 GVEALHSLHIWALTVAQPVLSVHIAIAQNTDAQAVLKTPASSRLQGKFH---FHTVTIQIE 355
          ||| |||||||:|::|:|||||||:||||:|     |:|||:::| :|    || | | :
Db   364 GVEEVHELHVWQLAGSRIATAHIKCEDPASTMQVAKT---IKDVFHNHGIHATTIQPE 419
```

```
Qy   356 DYSEDMKD----CQ-ACQ 368
          || |  :||    ||  ||
Db   420 FASVGSKSSVVPCELACR 437
```

RESULT 33
D84459
probable cation transport protein [imported] - Arabidopsis thaliana
C;Species: Arabidopsis thaliana (mouse-ear cress)
C;Date: 02-Feb-2001 #sequence_revision 02-Feb-2001 #text_change 16-Feb-2001
C;Accession: D84459
R;Lin, X.; Kaul, S.; Rounsley, S.D.; Shea, T.P.; Benito, M.I.; Town, C.D.; Fujii, C.Y
M.; Koo, H.; Moffat, K.S.; Cronin, L.A.; Shen, M.; VanAken, S.E.; Umayam, L.; Tallon,
euss, D.; Nierman, W.C.; White, O.; Eisen, J.A.; Salzberg, S.L.; Fraser, C.M.; Venter
Nature 402, 761-768, 1999
A;Title: Sequence and analysis of chromosome 2 of the plant Arabidopsis thaliana.
A;Reference number: A84420; MUID:20083487; PMID:10617197
A;Accession: D84459
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-300 <STO>
A;Cross-references: GB:AE002093; NID:g4895164; PIDN:AAD32753.1; GSPDB:GN00139
A;Gene: At2g04620
A;Map position: 2
C;Superfamily: zinc transporter ZnT-2

Query Match   17.4%;   Score 334; DB 2; Length 300;
Best Local Similarity   27.3%;   Pred. No. 8.9e-21;
Matches 79; Conservative 76; Mismatches 118; Indels 16; Gaps 6;

```
Qy    83 LFMIGEVVGGYLAHSLAVMTDAAHLLTDFASMLISLFSLWMSSRPATKTMNFGWQRAEIL 142
          |||||:|| ||||||||||||:||:|||||:|:||:|:|  :|  |  | || ||||:
Db    14 IITLSEVVGGLISGSLALLGDSLHNFSDTMSLLGSPIAMKISEKPKNKKYTFGYRSEII 73
```

```
Qy   143 GALVSVLSINWVTGVLVYLAVERLISGDYEIDGGTMLITSGCAVAVNIIMGLTLHOSGHG 202
          ||||:|:| ||:|:||:|||||| | |  || :|| |::||||||||||||||||:|||
Db    74 VAFLNSVSIFVVSTLVVIERAVKRLLS-PAIVHTSVILLVSSIGLAANFSVILLHT--- 128
```

```
Qy   203 HSHGTTNQQEENPSVRAAFIHVIGDFMQSMGVLVAAYILYFKPEYKYVDPICTFVFSILV 262
          |||| :||||| |   |||||:||||||:|||||| |: |  ||||||||:|:|:||| |
Db   129 HS------KESMNVRSAYLHLIADTLSSIIVVLGAVFMRVWKIY-WLDPVLAFVIALYM 180
```

```
Qy   263 LGTTLTILRDVILVLMEGTPKGVDFTAVRDLLLSVEGVEALHSLHIWALTVAQPVLSVHI 322
          |||||:||||||||||:|||||||||||||||||||||:|||||||||:|:|:|||||:|
Db   181 FKEAYEIVKRSLEILMEASP-NLDFEKIKEIEKIEGVRNAHHFHAWRVGEKEIHFECHV 239
```

```
Qy   323 AI--AQNTDAQAVLKTASSRLQGKHFHFHVTIQIEDYSEDMKDCQAC 367
          |:  |:  :|:| ||||: |: |  :| || :|  ||  ||
Db   240 EVDNMELKDAQKIIDEIEKRLK-KYGITHTTTVQLE-------CERC 277
```

```
Qy   69  KAQRQLYVASAICLLFMIGEVVGGYLAHSLAVMTDAAHLLTDFASMLISLFSLWMSSRPA  128
           |:|: ||| || ||||| ||:||  |||:| |:::||| |||:||||:|:|||||
Db   22  KKSRKIALFLLINTAYMWVEFVAGFMSNSLGIIISDACHMLFDCAALAIGLYASYISRLPA   81

Qy  129  TKTMNFGWQRAEILGALVSVLSIWVVTGVLIYVLVLAVERLISGDYEIDGGTMLITSGCAV  188
            |:!| |:::|| ||||| || ||| ||||||||:  ||||||||||
Db   82  NHQYNYGRGFEVLSGYVNAVFLVVLVGALIVLESIERILD-PQEISTNSLLVVSVGGLLV   140

Qy  189  NIIMGLTLHQSGHSHGHTTNQQEENPSVRAAFIHVIGDFMQOSMGVLVAAYILYFKPEYK   248
           || | |||| || ||| |:||| | |:||:: |:| | ||| :||||:|:::|
Db  141  NIV-GLIFFHEHHHAHGGS------GIFLHVLADTMGSGVVVISTLLIKYK-GWL       188

Qy  249  YVDPICTFVFSILVLGTTLTLILRDVILVLIMEGTPKG--VDFTAVRDLLLSVEGVEALHSL 306
           |: |||||| |:| |:::: ||||| || ||||||| |:|||:||||:| |||
Db  189  VADPASSIFISILIIASVIPILRNSAEILLQRVPRAHRQDLKEAMRNILKTKGVCSIQRL   248

Qy  307  HIWALTVAQPVLSVHIAIAQNTDAQAVLKTASSRLQGKFHFHTVTIQIE            355
           || |:||||||||||||||||||||||||:|| : || |||:||:
Db  249  HVWSFTNSDDVVATLHLLVSADSD-KTDTKLQVSRLLEDAGVKDWTLQVE           296
```

RESULT 34
S54302
zinc transporter ZnT-1 - mouse
C;Species: Mus musculus (house mouse)
C;Date: 15-Jul-1995 #sequence_revision 01-Sep-1995 #text_change 05-Nov-1999
C;Accession: S54302
R;Palmiter, R.D.; Findley, S.D.
EMBO J. 14, 639-649, 1995
A;Title: Cloning and functional characterization of a mammalian zinc transporter that co
A;Reference number: S54302; MUID:95188868; PMID:7882967
A;Accession: S54302
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-503 <PAL>
A;Cross-references: EMBL:U17132; NID:g577840; PIDN:AAA79233.1; PID:g577841
C;Genetics:
A;Genetics:
A;Introns: 203/1

Query Match   17.3%;   Score 333;   DB 2;   Length 503;
Best Local Similarity 23.7%; Pred. No. 2e-20;
Matches 103; Conservative 72; Mismatches 121; Indels 138; Gaps 12;

```
Qy   68  GKAQRQLYVASAICLLFMIGEVVGGYLAHSLAVMTDAAHLLTDFASMLISLFSLWMSSRP  127
           |:: || ||||| |||||||||||||||||||||||||||||||||||||||||||||
Db    5  GRNRGRLLCMLLLTFMFVLEVVVSRVTASLAMLSDSFHMLSDVLALLVVALVAERFARRT   64

Qy  128  -ATKTMNFGWQRAEILGALVSVLSIWVVTGVLVYLAVERLISGDYEIDGGTMLITSGCAV  186
            ||||||||||||| |||| || |||: :|:| ||||:||||||||||||||||||
Db   65  HATQKNTFGWIRAEVMGALVNAIFLTGLCFAILLEAVERFIE-PHEMQQPLVVLSVGVAG  123

Qy  187  AVNIIMGLTL--HQS-----------GHGHSHG-----------------           212
           |:|||| |  ||| :||| |
Db  124  LIVNVLGICLFHHHSGEGQGAGHGHSHGHGHLAKGARKAGRAGVEAGAPGRAPQEE      183

Qy  213  ENP----------------SVRAA                                     220
           :||               :|
Db  184  TNTLVANTSNSGLKADQAEPEKLRSDDPVDVQVNGNLIQESDNLEAEDNRAGQLNMRGV   243

Qy  221  FIHVIGDFMQSMGVLVAAYILYFK-------------PE---                   246
           |:||:|||:|:|||:|||||:|:||              |:
Db  244  FLHVLGDALGSVIVVNALVFYFNWKGCTEDFDFCTNPCFPDPCKSSVEIINSTQAPMRDA  303

Qy  247  ---YKVVDPICTFVFSILVLGTTLTLILRDVILVLIMEGTPKGVDFTAVRDLLLSVEGVEA 302
              |: |||||| |:| |:::: ||||| || ||||||||:|||:||||:| |||
Db  304  GPCWVLYLYDPFLCIMWCILLYTTVPLLKESAILLQTVPKQIDIKHLVKELRDVGVEE    363

Qy  303  LHSLHWALTVAQPVLSVHIAIAQNTDAQAVLKTASSRLQGKFH---FHTVTIQIEDYSE  359
           |||||| |||||||||||||||||||||||||||||:|| : || |||:||::::
Db  364  VHELHVWQLAGSRIIATAHICEDPASYMQVAKT----IKDVFHNHGIHATTIQPEPASV  419

Qy  360  DMKD----CQ-ACQ                                               368
            ::|     ||
Db  420  GSKSSVLPCELACR                                              433
```

RESULT 35
AG2540
cation efflux system protein [imported] - Nostoc sp. (strain PCC 7120) plasmid pCC712
C;Species: Nostoc sp.
A;Note: Nostoc sp. strain PCC 7120 is a synonym of Anabaena sp. strain PCC 7120
C;Date: 14-Dec-2001 #sequence_revision 14-Dec-2001 #text_change 30-Jun-2002
C;Accession: AG2540
R;Kaneko, T.; Nakamura, Y.; Wolk, C.P.; Kuritz, T.; Sasamoto, S.; Watanabe, A.; Iriguc
Nakazaki, N.; Shimpo, S.; Sugimoto, M.; Takazawa, M.; Yamada, M.; Yasuda, M.; Tabata
DNA Res. 8, 205-213, 2001
A;Title: Complete Genomic Sequence of the Filamentous Nitrogen-fixing Cyanobacterium
A;Reference number: AB1807; MUID:21595285; PMID:11759840
A;Accession: AG2540
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-304 <KUR>
A;Cross-references: GB:AP003602; PIDN:BAB77253.1; PID:g17134695; GSPDB:GN00181
C;Genetics:
A;Genetics:
A;Gene: all7610
A;Genome: plasmid
C;Superfamily: zinc transporter ZnT-2

Query Match   17.3%;   Score 332;   DB 2;   Length 304;
Best Local Similarity 25.2%; Pred. No. 1.3e-20;
Matches 78; Conservative 81; Mismatches 128; Indels 22; Gaps 6;

```
Qy   47  AQSNHHCHAQKGPDSHCDPKKGKAQRQLYVASAICLLFMIGEVVGGYLAHSLAVMTDAAH  106
            |:||:|:  | |: |||:|| ||||
Db    2  SHSHSHGHSHE-PINY------NRAFIISVALNTGFVVIEVVYGLAANSLALLADAGH     52

Qy  107  LLTDFASMLISLFSLWMSSRPATRKTMNGWQRAEILGALVSVLSIWVVTGVLVYLAVERL  166
           |:||||||| | ||| |:| ||||| | | ||||:|| || |||: :|:|
Db   53  NLSDVLGLLLAWGASFLGRRQPTPRRTYGLRRSSTLAALLNAVLVLVASGAVAWEAIQRF  112

Qy  167  ISGDYEIDGGTMLITSGCAVAVNIIMGLTLHQSGHGHSHGHSHGTTNQQEENPSVRAAFIHVIG 226
           ||||||||||||||||||  | :| || :|
Db  113  AQPS-PVSGSTIIGVAAVGIAINMGSAL-MFLSG------------RERDLNIRGAFIHLVA   160

Qy  227  DFMQOSMGVLVAAYILYFKPEYKVYDPICTFVFSILVLGTTLTLILRDVILVLIMEGTPKGVD  286
           || ::|||:||||||| |: ||| : | ::|| |:
Db  161  DAAVSLGAVLAG-IAIVAIGWLWFDPVVSLMIVVVIVVSTWQLFQESLNLVTDAVPAGIE    219

Qy  287  FTAVRDLLLSVEGVEALSHIWALTVAQPVLSVHIAIAQNTDAQAVLKTASSRLQGKFH     346
           ||||:|||| | || | :::| |||||||||||||||||||||||||||||||||
Db  220  PLAVRFTLAELPGVSSIHDLHIWAMSTFETALTPAHLIMPAGNLSDAFLVQVNQQLHDHFG  279

Qy  347  FHTVTIQIE                                                    355
           ||:|| ||
Db  280  IEHTTIHIE                                                   288
```

RESULT 36
T43140
probable heavy metal ion resistance protein - fission yeast (Schizosaccharomyces pomb
C;Species: Schizosaccharomyces pombe
C;Date: 11-Jan-2000 #sequence_revision 11-Jan-2000 #text_change 21-Jul-2000
C;Accession: T43140
R;Yoshioka, S.; Kato, K.; Nakai, K.; Okayama, H.; Nojima, H.
DNA Res. 4, 363-369, 1997
A;Title: Identification of open reading frames in Schizosaccharomyces pombe cDNAs.
A;Reference number: Z17323; MUID:98162722; PMID:9501991
A;Accession: T43140
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: mRNA
A;Residues: 1-387 <YOS>
A;Cross-references: EMBL:D89236; NID:g1749679; PIDN:BAA13897.1; PID:g1749680
A;Experimental source: strain PR745
C;Superfamily: zinc transporter ZnT-2

```
Query Match
Best Local Similarity   23.7%;   Pred. No. 4.9e-19;
Matches  88;  Conservative  74;  Mismatches 118;  Indels  92;  Gaps  10;

Qy   71 QRQLYVASAICLLFMIGEVVGGYLAHSLAVMTDAAHLLTDFASMLISLFSLWMSSRPATK 130
           |:|:| ||  :|| |  |  ||||| ||:|||||||||:|:|| ||:|: | : ||
Db    7 QTRIILLLGIDVTFFIEIIGYAIDSLALIADSFHMLNDIVSLLV---ALW-ATRLAHS  62

Qy  131 TMN-----FGWQRAEILGALVSVLSIWVTGVLVLYAVERLISGDYEIDGGTMLITSGCA 185
Db   63 TSHEPKYTYGWQRAEILGALSNGVFLIALCMFIFMEAIERFIEPP-SVSNPTLMFFVGSL 121

Qy  186 VAVNIMGLTL-HQSGHGHSHG----------- 212
Db  122 GLLSNFVGIFLFHDHGHDHPHTAQNYDFPEEDDIESVLPSTIVHRCNTSQQEVSHTHT 181

Qy  213 ----------------------------------ENPSVRAAFIHVIGDF 228
Db  182 QVADSATESSPLLSYTGNHNGAGTSKPVNNHGSIEQDAPKQTKKRNLNMHGVFLHVLGDA 241

Qy  229 MQSMGVLVAA-YILYFKPEYKYV-DPICTFVFSILVLGTTILTILRDVILVLMEGTPKGVD 286
Db  242 LGNIGVISAALFIKYTDYSWRFLFDPCISILLTFIILFSAIPLCKSAALILLQVAPQSIK 301

Qy  287 FTAVRDLLLSVEGVEALHSLHIWALTVAQPVLSVHIAIAQNTD--AQAVLKTASSRLQG 343
Db  302 LDDVSNLINHLDGVESVHELHIWQLSDVKLIATVHVCVTLPDDKGESYTKLTTDIRNVLQ 361

Qy  344 KFHFHTVTIQIE 355
Db  362 SFGIYDVTIQPE 373


RESULT 37
T38252
probable zinc/cadmium resistance protein - fission yeast (Schizosaccharomyces pombe)
C;Species: Schizosaccharomyces pombe
C;Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 17-Mar-2000
C;Accession: T38252
R;Brown, D.; Churcher, C.M.; Barrell, B.G.; Rajandream, M.A.; Wood, V.
submitted to the EMBL Data Library, August 1995
A;Reference number: Z21781
A;Accession: T38252
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-387 <BRO>
A;Cross-references: EMBL:Z98559; PIDN:CAB11166.1; GSPDB:GN00066; SPDB:SPAC23C11.14
A;Experimental source: strain 972h-; cosmid c23c11
C;Genetics:
A;Gene: SPDB:SPAC23C11.14
A;Map position: 1
C;Superfamily: zinc transporter ZnT-2

Query Match
Best Local Similarity   23.7%;   Pred. No. 6e-19;
Matches  88;  Conservative  73;  Mismatches 119;  Indels  92;  Gaps  10;

Qy   71 QRQLYVASAICLLFMIGEVVGGYLAHSLAVMTDAAHLLTDFASMLISLFSLWMSSRPATK 130
Db    7 QTRIILLLGIDVTFFIEIIGYAIDSLALIADSFHMLNDIVSLLV---ALW-ATRLAHS  62

Qy  131 TMN-----FGWQRAEILGALVSVLSIWVTGVLVLYAVERLISGDYEIDGGTMLITSGCA 185
Db   63 TSHEPKYTYGWQRAEILGALSNGVFLIALCMFIFMEAIERFIEPP-SVSNPTLMFFVGSL 121

Qy  186 VAVNIMGLTL-HQSGHGHSHG----------- 212
Db  122 GLLSNFVGIFLFHDHGHDHPHTAQNYDFPEEDDIESVLPSTIVHRCNTSQQEVSHTHT 181

Qy  213 ----------------------------------ENPSVRAAFIHVIGDF 228
Db  182 QVADSATESSPLLSYTGNHNGAGTSKPVNNHGSIEQDAPKQTKKRNLNMHGVFLHVLGDA 241

Qy  229 MQSMGVLVAA-YILYFKPEYKYV-DPICTFVFSILVLGTTILTILRDVILVLMEGTPKGVD 286
Db  242 LGNIGVISAALFIKYTDYSWRFLFDPCISILLTFIILFSAIPLCKSAALILLQVAPQSIK 301

Qy  287 FTAVRDLLLSVEGVEALHSLHIWALTVAQPVLSVHIAIAQNTD--AQAVLKTASSRLQG 343
Db  302 LDDVSNLINHLDGVESVHELHIWQLSDVKLIATVHVCVTLPDDKGESYTKLTTDIRNVLQ 361

Qy  344 KFHFHTVTIQIE 355
Db  362 SFGIYDVTIQPE 373


RESULT 38
E81321
probable cation transport protein Cj1163c [imported] - Campylobacter jejuni (strain N
C;Species: Campylobacter jejuni
C;Date: 31-Mar-2000 #sequence_revision 31-Mar-2000 #text_change 03-Jun-2002
C;Accession: E81321
R;Parkhill, J.; Wren, B.W.; Mungall, K.; Ketley, J.M.; Churcher, C.; Basham, D.; Chil
C.W.; Quail, M.A.; Rajandream, M.A.; Rutherford, K.M.; VanVliet, A.; Whitehead, S.; Bar
Nature 403, 665-668, 2000
A;Title: The genome sequence of the food-borne pathogen Campylobacter jejuni reveals
A;Reference number: A81250; MUID:20150912; PMID:10688204
A;Accession: E81321
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-316 <PAR>
A;Cross-references: GB:AL139077; GB:AL111168; NID:g6968444; PIDN:CAB73417.1; PID:g696
A;Experimental source: serotype O2, strain NCTC 11168
C;Genetics:
A;Gene: Cj1163c
C;Superfamily: zinc transporter ZnT-2

Query Match
Best Local Similarity   26.8%;   Pred. No. 1.3e-18;
Matches  84;  Conservative  76;  Mismatches 134;  Indels  20;  Gaps  7;

Qy   45 LAAQSNHHCHAQKGPD---SHCDPKKGKAQRQLYVASAICLLFMIGEVVGGYLAHSLAVM 101
Db   10 LANKSCHHNHEEHSHEHHHHSHAD-ARSVDKKLIKLISLLMTFSMMLVQFIYSILSNSLALL  68

Qy  102 TDAAHLLTDFASMLISLFSLWMSSRPATKTMNFGWQRAEILGALVSVLSIWVTGVLVYL 161
Db   69 SDTLHMFSDVFALALSFLAIIAVEKWQDHQKTFGYFRLEVLVAFINALTIILSALFIYE 128

Qy  162 AVERLISGDYEIDGGTMLITSGCAVAVNIIMGLTLHQSGHGHSHGTTNQQEENPSVRAAF 221
Db  129 AIEKFIN-PKEIDAKTMIIVAILGFLVNGTNALMFKGAN-------LENVNMKSAF 177

Qy  222 IHVIGDFEMQSMGVLVAAYILYFKPEYKYVDPICTFVFSILVLGTTILTILRDVILVLMEGT 281
Db  178 LHMMSDLLGSLAVIIGGIVVYFS-DIVYIDFILAIVLSILLLRWAIILLKQSANVLLESS 236

Qy  282 PKGVDFTAVRDLLLSVEGVEALHSLHIWALTVAQPVLSVH--IAIAQNTDAQAVLKTASS 339
Db  237 P--VDIEKVRQVLLLNPSVDEVVDLHITQITNKMLVASMHLKVRVCNLKEFEKLSQDLSH 294

Qy  340 RLQGKFHFHVTIQ 353
Db  295 KLLHEFEIGHITIQ 308


RESULT 39
H71078
probable cation efflux system protein czcD - Pyrococcus horikoshii
C;Species: Pyrococcus horikoshii
C;Date: 14-Aug-1998 #sequence_revision 14-Aug-1998 #text_change 20-Jun-2000
C;Accession: H71078
R;Kawarabayasi, Y.; Sawada, M.; Horikawa, H.; Haikawa, Y.; Hino, Y.; Yamamoto, S.; Se
M.; Ohfuku, Y.; Funahashi, T.; Tanaka, T.; Kudoh, Y.; Yamazaki, J.; Kushida, N.; Ogu
DNA Res. 5, 55-76, 1998
```

```
A;Title: Complete sequence and gene organization of the genome of a hyper-thermophilic a
A;Reference number: A71000; MUID:98344137; PMID:9679194
A;Accession: H71078
A;Status: preliminary; nucleic acid sequence not shown; translation not shown
A;Molecule type: DNA
A;Residues: 1-295 <KAW>
A;Cross-references: GB:AP000004; NID:g3261131; PIDN:BAA29990.1; PID:g3257307
A;Experimental source: strain OT3
A;Note: this accession replaces an interim accession for a sequence replaced by GenBank
A;Genetics:
A;Gene: PH0896
C;Superfamily: zinc transporter ZnT-2
```

```
Query Match             15.4%;   Score 296.5; DB 2;  Length 295;
Best Local Similarity   26.8%;   Pred. No. 1.3e-17;
Matches    77;   Conservative   74;   Mismatches  119;   Indels   17;   Gaps    7;

Qy    71 QRQLYVASAICLLFMIGEVGGGYLAHSLAVMTDAAHLLTDFASMLISLFSLWMSSRPATK 130
            :| :| | :|:: | | ::|:|:|  |  |:|: | | ||:|:|:|:|| ||
Db     8 KRKMMLSFSLNIGITLAEVIGGIISGSLALLSDSLHNFSDSMGIFASYFAIKIGERKKNE  67

Qy   131 TMNFGWQRAEILGALVSVLSIWVVTGVLVLVLAVERLISGDYEIDGGTMLITSGCAVAVNI 190
            | |   |: : ||| |: |:|| | :|:|| :|: |  | |:: |||  | | ||| |:
Db    68 KFTGYKRAEILVAFANSAVLVGVSLFIFEAYKREKSPE--PIKGSLMLFVVALIGLFANL 126

Qy   191 IMGLTLHQSGHGHSHGTTNQQEENPSVRAAFIHVIGDFMQSMGVLVAAYLYFKPEYKYV 250
            |:|| :|: :|::| | || ||::| |: ||| :| | :|: | | | :| |
Db   127 ASVILLHE---HAH------ESMNIRSAYLHLVSDSLSSIAVVLGG-IAITKWNVFWI 174

Qy   251 DPICTFVFSILVLGTTLTILRDVILVLMEGTPKGVDFTAVRDLLSVEGVEALHSLHIWA 310
            | | :|:  || ::|| |  |:|:|::|:|| : |||| || | | :|  |:|| ::
Db   175 DPLIVLISLYILHESYGILKESIEVLMEAAP-NLDFQKIKEEIESIPKVRNAHHFHAWR 233

Qy   311 LTVAQPVLSVHIAIAQN--TDAQAVLKTASSRLQGKFHFHTVTIQIE 355
            |||: |:||  |  |:  | | |: |:| |   | |  | |:||| :
Db   234 VGENEIHFECHVETEDMLLSEAQKLIDAIADKLK-KFGITHVTIQLE 279
```

```
RESULT 40
D33830
cation effux system membrane protein czcD - Alcaligenes eutrophus
C;Species: Alcaligenes eutrophus
C;Date: 10-Sep-1999 #sequence_revision 10-Sep-1999 #text_change 29-Oct-1999
C;Accession: D33830
R;Nies, D.H.; Nies, A.; Chu, L.; Silver, S.
Proc. Natl. Acad. Sci. U.S.A. 86, 7351-7355, 1989
A;Title: Expression and nucleotide sequence of a plasmid-determined divalent cation effl
A;Reference number: A33830; MUID:9001477; PMID:2678100
A;Accession: D33830
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-199 <NIE>
A;Cross-references: GB:M26073
C;Comment: This sequence appears to be homologous to PIR:H64810 and other members of tha
C;Superfamily: membrane protein czcD
C;Keywords: membrane protein
```

```
Query Match             12.8%;   Score 245.5; DB 1;  Length 199;
Best Local Similarity   32.5%;   Pred. No. 1.8e-13;
Matches    67;   Conservative   44;   Mismatches  80;   Indels   15;   Gaps    5;

Qy    61 SHCDPKKGKAQRQLYVASAICLLFMIGEVGGGYLAHSLAVMTDAAHLLTDFASMLISLFS 120
            | :|  |  :||:|:||| |  ||| |||| ||||||||||| :| | |||:|||:|:|
Db     6 SHDHP--GGNERSLKIALALIGTFLIAEVGGGVMTKSLALISDAAHMLTDIVALALALAA  63

Qy   121 LWMSSRPATKTMNFGWQRAEILGALVSVLSIWVVTGVLVLVLAVERLISGDYEIDGGTMLI 180
            |  ||||  | :||| |||||||||:|||||||||||:|: :|||||||||||||||| |
Db    64 IAIAKRPADKKRTFGYYRFELLAAAFNALLFGVAIYILYEAYLRLKSPP-QIESTGMFV 122

Qy   181 TSGCAVAVNIIMGLTLHQSGHGHSHGTTNQQEENPSVRAAFIHVIGDFMQSMGVLVAAYI 240
            :||| || | |||| :|:: :|| |:|: :: ::::      :     ||:: | | :|
Db   123 VAVLGLINLI-SMRMLSSG---------QSSSLNVKGAYLEVWSDLLGSVGVIAGAII 171
```

```
Qy   241 LYFKPEYKYVDPICTFVFSILVLGTT 266
            :  | :|| |:| |  || |:|:| |
Db   172 IRF-TGWAWVDSAIAVLIGLWVLPRT 196
```

```
Search completed: March 18, 2003, 15:41:28
Job time : 25 secs
```

OM protein - protein search, using sw model

GenCore version 5.1.4_p5_4578

Run on: March 18, 2003, 15:36:33 ;

Title:
Perfect score: 1923
Sequence: 1 MEAKEKQHLLDARPAIRSYT.........QIEDYSEDMKDCQACQGPSD 372
Scoring table: BLOSUM62
                 Gapop 10.0 , Gapext 0.5
Searched: US-09-691-219-2

Searched: 908470 seqs, 133250620 residues

Total number of hits satisfying chosen parameters: 908470

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
                 Maximum Match 100%
                 Listing first 120 summaries

Database :   A_Geneseq_101002:*

Search time 40 seconds
(without alignments)
1239.231 Million cell updates/sec

1: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1980.DAT:*
2: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1981.DAT:*
3: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1982.DAT:*
4: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1983.DAT:*
5: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1984.DAT:*
6: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1985.DAT:*
7: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1986.DAT:*
8: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1987.DAT:*
9: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1988.DAT:*
10: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1989.DAT:*
11: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1990.DAT:*
12: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1991.DAT:*
13: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1992.DAT:*
14: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1993.DAT:*
15: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1994.DAT:*
16: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1995.DAT:*
17: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1996.DAT:*
18: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1997.DAT:*
19: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1998.DAT:*
20: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1999.DAT:*
21: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA2000.DAT:*
22: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA2001.DAT:*
23: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA2002.DAT:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	1923	100.0	372	23	ABB83084	Human transporter
2	1923	100.0	372	23	AAE29905	Human transporter,
3	1493	77.6	358	23	ABB83085	Transporter relate
4	911.5	47.2	369	23	ABG60226	Human zinc transpo
5	907.5	46.9	320	22	ABG60224	Human zinc transpo
6	902.5	46.9	320	23	AAB60094	Human zinc transpo
7	902.5	46.9	320	23	AAU99907	Human transport pr
8	902.5	46.9	320	23	AAI17562	Human 84233 metal
9	878.5	45.7	472	22	ABB60324	Drosophila melanog
10	838.5	43.6	422	23	ABG60025	Human zinc transpo
11	811.5	42.2	1677	22	ABB69976	Drosophila melanog
12	722.5	37.6	429	23	ABG61911	Prostate cancer-as
13	722.5	37.6	438	23	AAE24062	Human prostate spe
14	656.5	34.1	474	22	AAE09322	Rice zinc transpor
15	656.5	33.9	448	22	AAE09321	Maize zinc transpo
16	649.5	33.8	398	22	AAG22263	Arabidopsis thalia
17	649.5	33.8	398	21	AAG43478	Arabidopsis thalia
18	643	33.4	344	22	AAG22264	Arabidopsis thalia
19	643	33.4	344	21	AAG43479	Arabidopsis thalia
20	613.5	31.9	349	22	AAE09323	Soybean zinc trans
21	608	31.6	330	22	AAG22265	Arabidopsis thalia
22	608	31.6	330	21	AAG43480	Arabidopsis thalia
23	565	29.4	452	22	ABB65511	Drosophila melanog
24	500.5	26.0	375	22	AAG31822	Arabidopsis thalia
25	500.5	26.0	385	21	AAG31821	Arabidopsis thalia
26	498.5	25.9	359	21	AAE09320	Maize zinc transpo
27	469	24.4	322	23	AAG31823	Arabidopsis thalia
28	469	24.4	322	22	AAU99931	Human 85041 cation
29	442.5	23.0	289	23	AAU99932	Human 84234 cation
30	430	22.4	144	23	ABB51142	Human ORF48 protei
31	414.5	21.6	303	23	ABB48454	Listeria monocytog
32	408.5	21.2	342	23	ABP33945	Staphylococcus epi
33	385	20.0	359	23	ABP40773	Staphylococcus epi
34	366	19.0	449	22	ABB57822	Drosophila melanog
35	366	19.0	449	22	ABB64435	Drosophila melanog
36	358	18.6	442	11	AAR04584	Protein product of
37	357	18.6	429	17	AAR95451	Yeast OSR. Saccha
38	350	18.2	318	22	AAG91156	C glutamicum prote
39	350	18.2	318	22	AAB76797	Corynebacterium gl
40	347.5	18.1	507	22	AAG67549	Amino acid sequenc
41	346	18.0	512	23	AAE13348	Human zinc transpo
42	342	17.8	157	23	AAY86241	Wheat zinc transpo
43	333.5	17.3	490	21	AAU11185	Human secreted pro
44	322	16.7	485	23	AAU99906	Human 83378 metal
45	322	16.7	520	23	ABP51103	Human MDDT SEQ ID
46	320	16.6	310	22	AAU61849	Propionibacterium
47	317.5	16.5	407	22	AAM41465	Human polypeptide
48	310	16.1	434	22	AAU87563	Novel central nerv
49	281.5	14.6	500	22	AAW74581	Novel central nerv
50	281.5	14.6	564	19	AAW74581	5' fragment of mem
51	281.5	14.6	594	22	AAW74581	Human protein sequ
52	281.5	14.6	378	22	AAM39679	Human zinc transpo
53	279.5	14.5	765	21	AAU11185	Human PROI879 poly
54	278.5	14.5	692	22	AAW94284	Rabbit membrane se
55	278.5	14.4	594	19	AAW74579	Human protein sequ
56	276.5	14.4	166	22	AAU99927	Human 83378 second
57	272.5	14.2	594	19	AAW74580	Human membrane pro
58	272.5	14.2	761	23	AAU99940	Human 85041 metal
59	269.5	14.0	366	22	ABB65584	Drosophila melanog
60	262	13.6	366	22	AAB38896	Human polypeptide
61	241	12.5	376	23	AAU99909	Human 84234 metal
62	241	12.5	388	22	AAM40682	Human polypeptide
63	240	12.5	376	23	AAB88256	Human polypeptide
64	238.5	12.4	377	23	AAE22907	Human transporter
65	234	12.2	72	23	ABP07382	Human ORFX protein
66	233	12.1	304	23	ABS54053	Lactococcus lactis
67	229.5	11.9	349	22	AAU69449	Human purified sec
68	229	11.9	300	21	AAY81582	Streptococcus pneu
69	218.5	11.4	287	23	ABP25799	Streptococcus poly
70	213.5	11.1	293	21	AAY86316	Human secreted pro
71	213.5	11.1	307	21	AAY86410	Human gene 27-enco
72	213.5	10.9	167	21	ABP38022	Staphylococcus epi
73	210.5	10.9	174	23	ABP38022	Staphylococcus epi
74	208	10.8	208	23	AAU99926	Human 83378 first
75	206.5	10.7	151	23	AAM40116	Human 83378 prote
76	206.5	10.5	373	22	AAP31895	Human polypeptide
77	201.5	10.5	375	22	AAU61324	Human polypeptide
78	201.5	10.3	327	22	AAM41302	Human polypeptide
79	197.5	10.1	513	23	ABP71353	Propionibacterium
80	194.5	9.8	291	23	ABP25800	Drosophila melanog
81	187.5	9.8	240	23	AAM52621	Streptococcus poly
82	186.5	9.7	130	23	AAW79301	Human zinc ion tra
83	186.5	9.7	461	23	AAU99908	A Staphylococcus a
82	181.5	9.8	130	23		6708 metal
83	168.5	8.8	370	21	AAB41624	Human ORFX ORF1388

#					Accession	Description
84	165.5	8.6	200	23	ABB89882	Human polypeptide
85	165.5	8.6	241	22	AAG74108	Human colon cancer
86	164.5	8.6	96	23	AAU99930	Human 64708 second
87	160.5	8.3	125	23	AAU99929	Human 64708 first
88	156	8.1	291	23	ABB49310	Listeria monocytog
89	156	8.1	428	19	AAW56671	Caenorhabditis ele
90	153.5	8.0	57	23	ABP07949	Human ORFX protein
91	149	7.7	289	23	ABB48288	Listeria monocytog
92	143.5	7.5	199	22	AAG66785	zinc transporter h
93	138.5	7.2	130	20	AAY12709	Human 5' EST secre
94	135	7.0	411	23	ABP28132	Streptococcus poly
95	132	6.9	394	22	AAU37907	Streptococcus pneu
96	129.5	6.7	114	21	AAB45078	Human secreted pro
97	129.5	6.7	238	22	ABB89829	Human polypeptide
98	128	6.7	165	19	AAW38723	Streptococcus pneu
99	127.5	6.6	411	23	ABP28133	Streptococcus poly
100	122	6.5	218	22	ABP31393	Human ORF366 prote
101	120	6.3	286	22	AAG92803	C glutamicum prote
102	118.5	6.2	425	23	ABB53614	Lactococcus lactis
103	111.5	5.8	67	19	AAW79302	A Staphylococcus a
104	110	5.7	804	22	ABB59021	Drosophila melanog
105	109.5	5.7	314	23	ABB54787	Lactococcus lactis
106	107.5	5.6	283	22	AAB96208	Putative P. abyssi
107	107.5	5.6	348	21	AAG29213	Arabidopsis thalia
108	107.5	5.6	434	21	AAG29212	Arabidopsis thalia
109	107.5	5.6	451	21	AAG29211	Arabidopsis thalia
110	106	5.6	376	22	AAU44934	Propionibacterium
111	104.5	5.4	515	23	AAU74635	Oestrogen-regulate
112	100.5	5.2	154	23	ABP38855	Staphylococcus epi
113	100	5.2	431	22	AAU74441	Oestrogen-regulate
114	100	5.2	803	22	ABB59443	Drosophila melanog
115	99.5	5.2	458	22	AAG89932	C glutamicum prote
116	98	5.1	297	22	AAB78949	C. glutamicum SRT
117	98	5.1	575	23	ABP27731	Streptococcus poly
118	97.5	5.1	102	15	AAR46087	zinc resistant lik
119	97.5	5.1	1193	23	ABB90860	Herbicidally activ
120	96.5	5.0	372	22	ABB56373	Non-endogenous hum

ALIGNMENTS

RESULT 1
ABB83084
ID   ABB83084 standard; Protein; 372 AA.
XX
AC   ABB83084;
XX
DT   02-OCT-2002 (first entry)
XX
DE   Human transporter protein.
XX
KW   Human; transporter protein; zinc transporter; pharmacogenomic analysis;
KW   diagnosis; drug screening; gene therapy; kidney; testis; heart; placenta;
KW   small intestine; liver; chromosome 1.
XX
OS   Homo sapiens.
XX
PN   WO200224910-A2.
XX
PD   28-MAR-2002.
XX
PF   20-SEP-2001; 2001WO-US29218.
XX
PR   20-SEP-2000; 2000US-234160P.
PR   19-OCT-2000; 2000US-0691219.
XX
PA   (PESE) PE CORP NY.
XX
PI   Wei M, Ketchum KA, Di Francesco V, Beasley EM;
XX
DR   WPI; 2002-404954/43.
DR   N-PSDB; ABN83946, ABN83947.
XX
PT   Novel human transporter proteins, related to zinc transporter
PT   subfamily, useful as model for developing human therapeutic targets and
PT   serves as target for human therapeutics -
XX
PS   Claim 1(a); Fig 2; 75pp; English.
XX
CC   The invention relates to an isolated human transporter protein that is
CC   related to the zinc transporter subfamily. Polynucleotides and
CC   polypeptides of the invention are useful for treating a disease or
CC   condition mediated by human transporter protein. The proteins also
CC   provide a target for diagnosing a disease or predisposition to disease
CC   mediated by the peptide, and in pharmacogenomic analysis. The peptides
CC   are also useful for treating a disorders characterised by absence of,
CC   inappropriate or unwanted expression of the protein. The nucleic acids
CC   are also useful in drug screening assays and as a target for treatment
CC   by the compounds identified through drug screening. The invention also
CC   provides vectors for gene therapy in patients with aberrant expression
CC   of the gene encoding the transporter protein. The gene of the invention
CC   has been found to be expressed in humans in the kidney, testis, heart,
CC   placenta, small intestine and liver. The gene has been localised to
CC   human chromosome 1. The current sequence represents the human transporter
CC   protein.
XX
SQ   Sequence 372 AA;

Query Match   100.0%; Score 1923; DB 23; Length 372;
Best Local Similarity   100.0%; Pred. No. 2.8e-198;
Matches 372; Conservative   0; Mismatches   0; Indels   0; Gaps   0;

```
Qy    1 MEAKEKQHLLDARPAIRSYTGSLWQEGAGWIPLPRPGLDLQAIELAAQSNHHCHAQKGPD 60
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1 MEAKEKQHLLDARPAIRSYTGSLWQEGAGWIPLPRPGLDLQAIELAAQSNHHCHAQKGPD 60

Qy   61 SHCDPKKGKAQRQLYVASAICLLFMIGEVVGGYLAHSLAVMTDAAHLLTDFASMLISLFS 120
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   61 SHCDPKKGKAQRQLYVASAICLLFMIGEVVGGYLAHSLAVMTDAAHLLTDFASMLISLFS 120

Qy  121 LWMSSRPATKTMNFGWQRAEILGALVSVLSIWVVTGVLVYLAVERLISGDYEIDGGTMLI 180
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  121 LWMSSRPATKTMNFGWQRAEILGALVSVLSIWVVTGVLVYLAVERLISGDYEIDGGTMLI 180

Qy  181 TSGCAVAVNIIMGLTLHQSGHGHSHGTTNQEENPSVRAAFIHVIGDEMQSMGVLVAAYI 240
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  181 TSGCAVAVNIIMGLTLHQSGHGHSHGTTNQEENPSVRAAFIHVIGDEMQSMGVLVAAYI 240

Qy  241 LYFKPEYKYVDPICTFVFSILVLGTTLTLIRDVILVLMEGTPKGVDFTAVRDLLLSVEGV 300
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  241 LYFKPEYKYVDPICTFVFSILVLGTTLTLIRDVILVLMEGTPKGVDFTAVRDLLLSVEGV 300

Qy  301 EALHSLHIWALTVAQPVLSVHIAIAQNTDAQAVLKTASSRLQGKFHFHTVTLQIEDYSED 360
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  301 EALHSLHIWALTVAQPVLSVHIAIAQNTDAQAVLKTASSRLQGKFHFHTVTLQIEDYSED 360

Qy  361 MKDCQACQGPSD 372
         ||||||||||||
Db  361 MKDCQACQGPSD 372
```

RESULT 2
AAE22905
ID   AAE22905 standard; Protein; 372 AA.
XX
AC   AAE22905;
XX
DT   09-AUG-2002 (first entry)
XX
DE   Human transporter and ion channel (TRICH) 4.
XX
KW   Human; transporter and ion channel; TRICH; transport disorder;
KW   diabetes mellitus; angina; Alzheimer's disease; neurological; epilepsy;
KW   stroke; Huntington's disease; meningitis; muscle; myocarditis; cancer;
KW   infectious myositis; arrhythmia; asthma; immunological; gene therapy;

```
KW   acquired immunodeficiency syndrome; AIDS; allergy; atherosclerosis;
KW   cell proliferative disorder; cerebroprotective; cirrhosis; hepatitis;
KW   transgenic; neuroprotective; anticonvulsant; nootropic; cytostatic;
KW   antiinflammatory; hepatotropic; psoriasis.
XX
OS   Homo sapiens.
XX
FH   Key            Location/Qualifiers
FT   Domain         141..159
FT                  /label= Transmembrane_domain
XX
PN   WO20022226684-A2.
XX
PD   21-MAR-2002.
XX
PF   14-SEP-2001; 2001WO-US28938.
XX
PR   15-SEP-2000; 2000US-232685P.
PR   22-SEP-2000; 2000US-234842P.
PR   29-SEP-2000; 2000US-236882P.
PR   05-OCT-2000; 2000US-239057P.
PR   13-OCT-2000; 2000US-240540P.
PR   18-OCT-2000; 2000US-241700P.
XX
PA   (INCY-) INCYTE GENOMICS INC.
XX
PI   Lee EA, Yue H, Lal PG, Walia NK, Baughn MR, Warren BA, Lee S;
PI   Sanjanwala MS, Yao MG, Ramkumar J, Thornton M, Gandhi AR;
PI   Policky JL, Elliott VS, Arvizu C, Raumann BE, Bruns CM, Naini A;
PI   Hafalia AJA, Nguyen DB, Xu Y, Lu DAM, Ison CH, Griffin JA;
PI   Reddy RM, Burford N;
XX
DR   WPI; 2002-333948/42.
DR   N-PSDB; AAD36301.
XX
PT   Polypeptides of human transporters and ion channels, useful for
PT   diagnosing, treating or preventing transport, neurological, muscle,
PT   immunological and cell proliferative disorders -
XX
PS   Claim 1; Page 144-145; 204pp; English.
XX
CC   The invention relates to human transporters and ion channels (TRICH)
CC   and their corresponding nucleic acid sequences. TRICH is useful for
CC   screening an agonist/antagonist that modulates its activity. TRICH is
CC   useful as an immunogen for preparing antibodies which are useful for
CC   diagnosing a condition associated with its expression in a
CC   subject, and for detecting and purifying it from a sample. TRICH DNA
CC   is useful as probe or a primer for assessing toxicity of a test
CC   compound. Composition comprising TRICH or its agonist is useful for
CC   treating a disease or condition associated with decreased expression
CC   of functional TRICH and composition comprising TRICH antagonist is
CC   useful for treating a disease or condition associated with TRICH
CC   overexpression of TRICH. TRICH sequence is used in the diagnosis and
CC   treatment of transport disorder e.g. diabetes mellitus, angina,
CC   Alzheimer's disease; neurological disorder e.g. epilepsy, stroke,
CC   Huntington's disease, infectious myositis, arrhythmias, asthma,
CC   e.g. myocarditis, bacterial and viral meningitis, muscle disorder
CC   immunological disorder e.g. acquired immunodeficiency syndrome (AIDS),
CC   allergies, atherosclerosis; and cell proliferative disorders e.g.
CC   cirrhosis, hepatitis, psoriasis and cancers. TRICH DNA is used in
CC   gene therapy. TRICH DNA is useful for creating knockin humanised
CC   animals (pigs) or transgenic animals (mice or rats) to model human
CC   disease. The present sequence is human TRICH protein.
XX
SQ   Sequence      372 AA;

Query Match           100.0%;
Best Local Similarity 100.0%;
Matches  372;    Conservative 0;    Mismatches 0;    Indels 0;    Gaps 0;
Score 1923;   DB 23;   Length 372;   Pred. No. 2.8e-198;

Qy    1 MEAKEKQHLLDARPAIRSYTGSLWQEGAGWIPLPRPGLDLQAIELAAQSNHHCHAQKGPD 60
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1 MEAKEKQHLLDARPAIRSYTGSLWQEGAGWIPLPRPGLDLQAIELAAQSNHHCHAQKGPD 60

Qy   61 SHCDPKKGKAQRQLYVASAICLLFMIGEVVGGYLAHSLAVMTDAAHLLTDFASMLISLFS 120
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   61 SHCDPKKGKAQRQLYVASAICLLFMIGEVVGGYLAHSLAVMTDAAHLLTDFASMLISLFS 120

Qy  121 LWMSSRPATKTMNFGWQRAEILGALVSVLSIWVVTGVLVYLAVEBRLISGDYEIDGGTMLI 180
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  121 LWMSSRPATKTMNFGWQRAEILGALVSVLSIWVVTGVLVYLAVEBRLISGDYEIDGGTMLI 180

Qy  181 TSGCAVAVNIIMGLTLHQSGHGHSHGTTMNQQEENPSVRAAFIHVIGDFMQSMGVLVAAYI 240
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  181 TSGCAVAVNIIMGLTLHQSGHGHSHGTTMNQQEENPSVRAAFIHVIGDFMQSMGVLVAAYI 240

Qy  241 LYFKPEYKYVDPICTFVFSILVLGTTLTIILRDVILVLMEGTPKGVDFTAVRDLLLSVEGV 300
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  241 LYFKPEYKYVDPICTFVFSILVLGTTLTIILRDVILVLMEGTPKGVDFTAVRDLLLSVEGV 300

Qy  301 EALHSLHIWALTVAQPVLSVHIAIAQNTDAQAVLKTASSRLQGKFHFHTVTIQIEDYSED 360
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  301 EALHSLHIWALTVAQPVLSVHIAIAQNTDAQAVLKTASSRLQGKFHFHTVTIQIEDYSED 360

Qy  361 MKDCQACQGPSD 372
         ||||||||||||
Db  361 MKDCQACQGPSD 372


RESULT 3
ABB83085
ID   ABB83085 standard; protein; 358 AA.
XX
AC   ABB83085;
XX
DT   02-OCT-2002 (first entry)
XX
DE   Transporter related protein.
XX
KW   Rat; transporter protein; zinc transporter; pharmacogenomic analysis;
KW   diagnosis; drug screening; gene therapy.
XX
OS   Rattus norvegicus.
XX
PN   WO20022224910-A2.
XX
PD   28-MAR-2002.
XX
PF   20-SEP-2001; 2001WO-US29218.
XX
PR   20-SEP-2000; 2000US-234160P.
PR   19-OCT-2000; 2000US-0691219.
XX
PA   (PENK-) PE CORP NY.
XX
PI   Wei M, Ketchum KA, Di Francesco V, Beasley EM;
XX
DR   WPI; 2002-404954/43.
XX
PT   Novel human transporter proteins, related to zinc transporter
PT   subfamily, useful as model for developing human therapeutics -
XX
PS   Disclosure; Page 75; 75pp; English.
XX
CC   The invention relates to an isolated human transporter protein that is
CC   related to the zinc transporter subfamily. Polynucleotides and
CC   polypeptides of the invention are useful for treating a disease or
CC   condition mediated by human transporter protein. The proteins also
CC   provide a target for diagnosing a disease or predisposition to disease
CC   mediated by the peptide, and in pharmacogenomic analysis. The peptides
CC   are also useful for treating a disorders characterised by absence of,
CC   inappropriate or unwanted expression of the protein. The nucleic acids
CC   are also useful in drug screening assays and as a target for treatment
CC   by the compounds identified through drug screening. The invention also
CC   provides vectors for gene therapy in patients with aberrant expression
```

```
CC   of the gene encoding the transporter protein. The gene of the invention
CC   has been found to be expressed in humans in the kidney, testis, heart,
CC   placenta, small intestine and liver. The current sequence represents a
CC   rat amino acid sequence that is relative to the transporter protein of
CC   the invention.
XX
SQ   Sequence   358 AA;

Query Match             77.6%;  Score 1493;  DB 23;  Length 358;
Best Local Similarity   80.1%;  Pred. No. 5.8e-152;
Matches 290;  Conservative 35;  Mismatches 29;  Indels  8;  Gaps  2;

Qy      15 AIRSYTGSLWQEGAGWIPLPRPGLDLQAIELAAQSNHHCHAQKGPDSHCDPKKGKAQRQL 74
            ||:|| ::||||   :  :  || :| |||||||||||||||||||||||||| : ::|
Db       1 ASRSFFGALWKSEASRI----PPVNLPSVELAVQSNHYCHAQKDSGSHPNSEKQRARRKL 56

Qy      75 YVASAICLLFMIGEVVGGYLAHSLAVMTDAAHLLTDFASMLISLFSLWMSSRPATKTMNF 134
            ||||||||| |||:|||::||:|||||||||||||||||||||||| :|||||||||||
Db      57 YVASAICLVFMIGEIIGGYLAQSLAIMTDAAHLLTDFASMLISLFSLWVSSRPATKTMNF 116

Qy     135 GWQRAEILGALVSVLSIWVVTGVLVYLAVERLISGDYEIDGGTMLITSGCAVAVNIIMGL 194
            |||||||||||||||||||||||||||| |||:|||||||||||||||||||||||||||
Db     117 GWQRAEILGALLSVLSIWVVTGVLVYLAVQRLISGDYEIKGDTMLITSGCAVAVNIIMGL 176

Qy     195 TLHQSGHGHSHG-----TTNQQEENPSVRAAFIHVIGDFMQSMGVLVAAYILYFKPEYKYV 250
            |||||||||||     :||:|||:||||||||||||||||:||||||||||||||||||||
Db     177 ALHQSGHGHSHGHSHEDSSQQQNPSVRAAFIHVVGDLLQSSVGVLVAAYILYFKPEYKYV 236

Qy     251 DPICTFVFSILVLGTTGTTLILRDVILVLMEGTPKGVDFTAVRDLLSVEGVEALHSLHIWA 310
            ||||| ||||||||||||:||:||||||||||||||| | |: || ||| |:||||||||
Db     237 DPICFLFSILVLGTTGTTLILRDVILVLMEGTPKGVDFTTVKNLLSVDGVEALHSLHIWA 296

Qy     311 LTVAQPVLSVHIAIAQNTDAQAVLKTPASSRLQGKFHFHTVTIQIEDYSEDMKDCQACQGP 370
            |||||||||||||||||||:||||:|||||||||:| :|||||||:|:|||||| |||||
Db     297 LTVAQPVLSVHIAIAQNVDAQAVLKVARDRLQGKFNFHTMVTIQIESYSEDMKSCQECQGP 356

Qy     371 SD 372
            |:
Db     357 SE 358
```

```
RESULT 4
ABG60226
ID   ABG60226 standard; Protein; 369 AA.
XX
AC   ABG60226;
XX
DT   30-JUL-2002  (first entry)
XX
DE   Human Zinc transporter-like protein NOV2c.
XX
PN   WO20024733-A2.
XX
PD   28-MAR-2002.
XX
PF   17-SEP-2001; 2001WO-US29115.
XX
PR   15-SEP-2000; 2000US-232675P.
PR   15-SEP-2000; 2000US-232676P.
PR   15-SEP-2000; 2000US-232679P.
```

```
PR   18-SEP-2000; 2000US-233382P.
PR   18-SEP-2000; 2000US-233402P.
PR   19-SEP-2000; 2000US-233521P.
PR   19-SEP-2000; 2000US-233522P.
PR   19-SEP-2000; 2000US-233801P.
PR   20-SEP-2000; 2000US-233960P.
PR   06-OCT-2000; 2000US-238398P.
PR   13-OCT-2000; 2000US-240284P.
PR   13-OCT-2000; 2000US-240498P.
PR   11-JAN-2001; 2000US-260973P.
PR   26-JAN-2001; 2001US-264274P.
PR   09-MAR-2001; 2001US-274862P.
XX
PA   (CURA-) CURAGEN CORP.
XX
PI   Mishra VS,  Syptek KA,  Taupier RJ,  Vernet CAM,  Colman SD,  Gorman L;
PI   Tcherney VT,  Malyankar UM,  Shenoy S,  Tcherney VT,  Padigaru M;
PI   Patturajan M,  Burgess CE,  Smithson G,  Millet I,  Peyman JA;
PI   Stone D,  Gunther E,  Ellerman K;
XX
DR   WPI; 2002-383182/41.
DR   N-PSDB; ABK71914.
XX
PT   New cytoplasmic, nuclear, membrane bound and secreted NOVX
PT   polypeptides, useful for treating cancers and tumours, lung disorders,
PT   haematopoietic disorders, autoimmune diseases and immune disorders -
XX
PS   Claim 1 ; Page 22; 210pp; English.

CC   The invention relates to an isolated NOVX polypeptide selected from
CC   NOV1a, NOV1b, NOV1ac, NOV2a, NOV2b, NOV2c, NOV3a, NOV3b, NOV4a, NOV4b,
CC   NOV5a, NOV5b or NOV6-NOV9 polypeptides, their mature form or variant.
CC   Also included are a nucleic acid encoding a NOVX protein or variant;
CC   a vector comprising the nucleic acid; a cell comprising the vector;
CC   an anti-NOVX antibody; and identifying agents that modulate the
CC   expression or activity of NOVX, the nucleic acid, antibody and
CC   modulators are useful in the diagnosis, treatment or prevention of
CC   developmental disorders, endocrine disorders, vascular disorders,
CC   infectious disease, anorexia, cancer, neurodegenerative disorders (e.g.
CC   Alzheimer's disease, Parkinson's disease, Huntington's disease, multiple
CC   sclerosis and amyotropic lateral sclerosis), acute brain injury (e.g.
CC   stroke, head injury and cerebral palsy), central nervous system disorders
CC   (e.g. depression, epilepsy and schizophrenia), lung disorders
CC   reproductive disorders, disorders affecting carbohydrate metabolism (e.g.
CC   galactosaemia and hereditary fructose intolerance), tissue disorders
CC   (e.g. Wiskott-Aldrich syndrome, thrombocytopaenia, night blindness and
CC   Pick's disease), disorders linked to abnormal angiogenesis, asthma,
CC   azoospermia, learning disabilities, facial dysmorphism, autoimmune
CC   encephalomyelitis, X-linked severe combined immunodeficiency, seizures,
CC   migraines, inflammation, autoimmune disorders, disorders affecting sleep,
CC   appetite, thermoregulation, pain, perception, hormone secretion and
CC   sexual behaviour, immune disorders, haematopoietic disorders or other
CC   disorders related to cell signal processing and metabolic pathway
CC   modulation, gastrointestinal diseases, respiratory disorders, blood
CC   disorders, hepatitis, trauma, regeneration, viral, bacterial or parasitic
CC   infections, hyper- or hypo-thyroidism, endometriosis, fertility,
CC   hypertension, arteriosclerosis, ischaemia, haemolytic anaemia, Werner
CC   syndrome, rheumatoid arthritis, Grave's disease, wound healing, X-linked
CC   mental retardation, psychotic and neurological disorders and neuronal
CC   degeneration. The present sequence represents a NOVX protein.
XX
SQ   Sequence   369 AA;

Query Match             47.4%;  Score 911.5;  DB 23;  Length 369;
Best Local Similarity   51.6%;  Pred. No. 2.8e-89;
Matches 176;  Conservative 63;  Mismatches 93;  Indels  9;  Gaps  2;

Qy      32 PLPRPGLDLQAIELAAQSNHHCHAQKGPDSHCDPKKGKAQRQLYVASAICLLFMIGEVVG 91
            | || ||| |   ||||||||||||||||||||||||||| :|||||||||||||||||
Db      38 PRERPE-----ELESGGMYHCHSGSKPTEKGANEYAYAKWELCSASAICFIPMIAEVVG 91

Qy      92 GYLAHSLAVMTDAAHLLTDFASMLISLFSLWMSSRPATKTMNEGWQRAEILGALVSVLSI 151
            ||||||||||||||||||||||||||||||||| : ||||||||  :||  ||||||||
```

```
Db    92   GHIAGSLAVVTDAAHLILDLTSLLLSFSLWLSSKPPSKRLTFGWHRAEILGALLSILCI 151

Qy   152   WVVTGVLVYLAVERLISGDYEIDGGTMLITSGCAVAVNIIMGLTLHQSGHGHSHGTTNQQ 211
           |||||||||||| :|| :|||:|  :||:|
Db   152   WVVTGVLVYLACERLLYPDYQIQATVMIIVSSCAVAANIVLTVVLHQRCLGHNH---KEV 208

Qy   212   EENPSVRAAFIHVIGDFMQSMGVLVAAYILYFKPEYKYVDPICFVFSILVLGTTLTILR 271
           :  |||||||||||:||:|| :|| |:|||||| |  |||| |||||:||||||||:||||
Db   209   QANASVRAAFVHALGDLFQSISVLISALIIYFKPEYKIADPICTFIFSILVLASTTTILK 268

Qy   272   DVILVLMEGTPKGVDFTAVRDLLLSVEGVEALHSIHWALTVAQPVLSVHIAIAQNTDAQ 331
           | :|||||| ||||||||||||:||||||||||||||||| |:|||| |:| | |:|
Db   269   DFSILLMEGVPKSLNYSGVKELILAVDGVLSVHSLHIWSLTMNQVILSAHVATAASRDSQ 328

Qy   332   AVLKTASSRLQGKFHFHTVTIQIEDYSEDMKDCQACQGPSD 372
           : :|||||  ||  | :| ||:|||:| : |||:| |
Db   329   VVRREIAKALSKSFTMHSLTIQMESPVDQDPDCLFCEDPCD 369
```

```
Query Match             47.2%;  Score 907.5;  DB 23;  Length 369;
Best Local Similarity   51.6%;  Pred. No. 7.5e-89;
Matches  176;  Conservative   62;  Mismatches   94;  Indels   9;  Gaps   2;
```

```
Qy    32   PLPRPGLDLQAIELAAQSNHHCHAQKGPDSHCDPRKGKAQRQLYVASAICLLFMIGEVVG  91
           || :||  |  : ||          |  |:  ||||||||||||||||||||||||||
Db    38   PRERPE------ELESGGMVHCHSGSKPTEKGANEYAYAKWKLCSASAICFFMIAEVVG  91

Qy    92   GYLAHSLAVMTDAAHLLTDFASMLISLFSLWMSSRPATKTMNFGWQRAEILGALVSVLSI 151
           |:|| |||||:|||||| ||:| ||:||||| | | |  |||  |||||||||:|:|| |
Db    92   GHIAGSLAVVTDAAHLILDLTSLLLSFSLWLSSKPPSKRLTFGWHRAEILGALLSILCI  151

Qy   152   WVVTGVLVYLAVERLISGDYEIDGGTMLITSGCAVAVNIIMGLTLHQSGHGHSHGTTNQQ 211
           |||||||||| :|| :|||:|  :||:|
Db   152   WVVTGVLVYLACERLLYPDYQIQATVMIIVSSCAVAANIVLTVVLHQRCLGHNH---KEV 208

Qy   212   EENPSVRAAFIHVIGDFMQSMGVLVAAYILYFKPEYKYVDPICFVFSILVLGTTLTILR 271
           :  |||||||||||:||:|| :|| |:|||||| |  |||| |||||:||||||||:||||
Db   209   QANASVRAAFVHALGDLFQSISVLISALIIYFKPEYKIADPICTFIFSILVLASTTTILK 268

Qy   272   DVILVLMEGTPKGVDFTAVRDLLLSVEGVEALHSIHWALTVAQPVLSVHIAIAQNTDAQ 331
           | :|||||| ||||||||||||:||||||||||||||||| |:|||| |:| | |:|
Db   269   DFSILLMEGVPKSLNYSGVKELILAVDGVLSVHSLHIWSLTMNQVILSAHVATAASWDSQ 328

Qy   332   AVLKTASSRLQGKFHFHTVTIQIEDYSEDMKDCQACQGPSD 372
           : :|||||  ||  | :| ||:|||:| : |||:| |
Db   329   VVRREIAKALSKSFTMHSLTIQMESPVDQDPDCLFCEDPCD 369
```

```
XX
AC   AAB60094;
XX
DT   28-MAR-2001 (first entry)
XX
DE   Human transport protein TPPT-14.
XX
KW   Human; transport protein; TPPT; transport disorder; metabolic disorder;
KW   neurological disorder; cardiovascular disorder; reproductive disorder;
KW   immune disorder; cancer.
XX
OS   Homo sapiens.
XX
PN   WO20007895 3-A2.
XX
PD   28-DEC-2000.
XX
PF   16-JUN-2000; 2000WO-US16668.
XX
PR   17-JUN-1999; 99US-0139923.
PR   10-AUG-1999; 99US-0148177.
PR   18-AUG-1999; 99US-0149357.
PR   28-OCT-1999; 99US-0162287.
XX
PA   (INCY-) INCYTE GENOMICS INC.
XX
PI   Lal P, Yang J, Yue H, Hillman JL, Tang YT, Bandman O, Burford N;
PI   Baughn MR, Azimzai Y, Lu DAM, Au-Young J, Patterson C;
XX
DR   WPI; 2001-041424/05.
DR   N-PSDB; AAF27714.
XX
PT   Isolated polypeptide with a human transport protein sequence is useful
PT   for the diagnosis, prevention and treatment of disorders associated
PT   with the immune, reproductive and cardiovascular systems -
XX
PS   Claim 2; Page 116-117; 165pp; English.
XX
CC   The present invention provides the protein and coding sequences for 43
CC   novel human transport proteins (designated TPPTs). These can be used in
CC   the diagnosis and treatment of transport, metabolic, neurological,
CC   reproductive, cardiovascular and immune disorders, and cell proliferative
CC   disorders such as cancer.
XX
SQ   Sequence   320 AA;

Query Match             46.9%;  Score 902.5; DB 22;  Length 320;
Best Local Similarity   53.1%;  Pred. No. 2.1e-88;
Matches 171;  Conservative 61;  Mismatches 87;  Indels   3;  Gaps   1;

Qy   51  HHCHAQKGPDSHCDPKKGKAQRQLYVASAICLLFMIGEVVGGYLAHSLAVMTDAAHLLTD 110
         :||||            |: ||    ||:||| ||| :|||||||||||:||||||||:||
Db    2  YHCHSGSKPTEKGANEYAYARWKLCSASAICFIFMIAEVVGGHIAGSLAVVTDAAHLLID 61

Qy  111  FASMLISLFSLMMSSRPATKTMNFGWQRAEILGALVSVLSIWVVTGVLIVYLAVERLISGD 170
         | |:||||||||:||:| : ||||||||||||||||||||||||||||||||||:||:|
Db   62  LTSFLLSLFSLWLSSKPPSKRLTFGWHRAEILGALLSILCIWVVTGVLVLVLACERLLYPD 121

Qy  171  YEIDGGTMLITSGCAVAVNIMGLLHQSGHGHSHGTTNQEENPSVRAAFIHVIGDFMQ 230
         |:||||| |:|||||||||||::|||  :|||||| ||||:|||||:|||||:|||
Db  122  YQIQATVMIIVSSCAVAANIVLTVTVVLLHQRCLGHNH---KEVQANASVRAAFVHALGDLFQ 178

Qy  231  SMGVLVAAYILYFKPEYKYVDPICTFVFSILVLGTTLTILRDVILVLMEGTPKGVDFTAV 290
         | :||||||||||   |:   |:| |||:||| ||: ||| |::| |: ||   ||
Db  179  SISVLISALIIYFKPEYKIADPICTFIFSILVLASTTILKDFSILLMEGVPKSLNYSGV 238

Qy  291  RDLLLSVEGYEALHSHINIWALTVAQPVLSVHIAIAQNTDAQAVLKTASSRLQGKFHFHTV 350
         : || : || :|:||||||||||: :||| ||:|:|| ||||:||:||||:: |||||
Db  239  KELILAVDGVLSVHSLHIWSLTMNQVLSAHVATAASRDSQVRREIAKALSKSFTMHSL 298

Qy  351  TIQIEDYSEDMKDCQACQGPSD 372
         ||||      || || ||| |
Db  299  TIQMESPVDQDPDCLFCEDPCD 320
```

```
RESULT 7
AAU99907
ID   AAU99907 standard; Protein; 320 AA.
XX
AC   AAU99907;
XX
DT   07-OCT-2002 (first entry)
XX
DE   Human 84433 metal transporter protein.
XX
KW   84233; metal transporter; human; infection;
KW   haematopoeitic disorder; blood clotting disorder; cancer;
KW   autoimmune disorder; leukaemia; immunological disorder;
KW   cardiovascular disorder; neurological disorder; cellular proliferation;
KW   red blood cell disorder; viral disease; neurological disorder.
XX
OS   Homo sapiens.
XX
FH   Key              Location/Qualifiers
FT   Domain           25..310
FT                    /note= "cation efflux domain"
XX
PN   WO200240656-A2.
XX
PD   23-MAY-2002.
XX
PF   14-NOV-2001; 2001WO-US045291.
XX
PR   14-NOV-2000; 2000US-248331P.
PR   14-NOV-2000; 2000US-248362P.
PR   14-NOV-2000; 2000US-248365P.
PR   30-NOV-2000; 2000US-250077P.
PR   30-NOV-2000; 2000US-250176P.
PR   30-NOV-2000; 2000US-250327P.
XX
PA   (MILL-) MILLENNIUM PHARM INC.
XX
PI   Meyers RE, Curtis RAJ, Glucksmann MA;
XX
DR   WPI; 2002-508325/54.
DR   N-PSDB; ABK88011.
XX
PT   Isolated 47476, 67210, 49875, 46842, 33201, 83378, 84233, 64708, 85041
PT   or 84234 polypeptides, useful as reagents or targets for treating or
PT   diagnosing pain or metabolic, liver, kidney, or cardiovascular
PT   disorders -
XX
PS   Claim 5; Page 238; 298pp; English.
XX
CC   This invention relates to the DNA and protein sequences of novel
CC   isolated 47476, 67210, 49875, 46842, 33201, 83378, 84233, 64708, 85041
CC   or 84234 proteins. The method of the invention is useful for treating a
CC   disorder characterised by aberrant activity of 47476, 67210, 49875,
CC   46842, 33201, 83378, 84233, 64708, 85041 or 84234-expressing cell in a
CC   subject. The protein molecules can act as novel diagnostic targets and
CC   therapeutic agents for controlling aberrant or deficient signal
CC   transduction resulting, in e.g., haematopoeitic disorders, including
CC   blood clotting disorders, autoimmune disorders, or disorders related to
CC   an inability to clear infections (e.g., viral or bacterial infections),
CC   as well as disorders related to abnormal cellular proliferation or
CC   differentiation, e.g., leukaemia. They may also be used to control
CC   disorders of metabolic imbalance (e.g., disorders of lipopolysaccharide
CC   biosynthesis or glycogen synthesis) immunological disorders
CC   cardiovascular disorders, neurological disorders, or cellular
CC   proliferation and/or differentiation disorders, e.g., cancer, cell
CC   motility and adhesion disorders differentiative disorders (e.g., brain
CC   cell disorders, viral diseases, neurological disorders, red blood
CC   disorders), pain or metabolic disorders, liver disorders, kidney
CC   disorders of the small intestine, disorders of metal ion
CC   imbalance, protein trafficking disorders and disorders associated with
CC   bone metabolism. The sequences of the invention are also useful for
```

```
CC   screening assays, predictive medicine (e.g., diagnostic assays,
CC   prognostic assays, monitoring clinical trials, and pharmacogenetics);
CC   and methods of treatment (e.g., therapeutic and prophylactic). The
CC   present sequence represents the human 84233 metal transporter protein
CC   of the invention.
XX
SQ   Sequence   320 AA;

Query Match   46.9%;  Score 902.5; DB 23;  Length 320;
Best Local Similarity   53.1%;  Pred. No. 2.1e-88;
Matches  171;  Conservative   61; Mismatches   87;  Indels    3;  Gaps    1;

Qy    51 HHCHAQKGPDSHCDPKKGKAQRQLYVASAICLLFMIGEVVGGYLAHSLAVMTDAAHLLTD 110
          |:||:|                | |||     ||::|||:| |||||::||||||||||
Db     2 YHCHSGSKPTEKGANEYAYAKWKLCSASAICFIFMIAEVVGGHIAGSLAVVTDAAHLLID 61

Qy   111 FASMLISLFSLWMSSRPATKTMNFGWQRAEILGALVSVLSIWVVTGVLVYLAVERLISGD 170
          |||  |||||| ||:||  || |||||||||||:||| ||||||||| |  |:|||:||
Db    62 LTSFLLSLFSIWLSSKPPSKRLTFGWHRAEILGALLSILCIWVVTGVLVYLACERLLYPD 121

Qy   171 YEIDGGTMLITSGCAVANIIMGLTLHQSGHGHSHGTTNQQEENPSVRAAFIHVIGDFMQ 230
          |:|| | | ||| ::||  ||| |||| ||||||:|||     | :| |||||||||
Db   122 YQIQATVMIIVSSCAVANIVLTVVLHQRCLGHNH--KEVQANASVRAAFVHALGDLFQ 178

Qy   231 SMGVLVAAYILYFKPEYKYVDPICTFVFSILVLGTTLTILTLRDVILVLMEGTPKGVDFTAV 290
          :|||| ||| | |||:| |||||||||||||:|:|| | ||||||| :||| ||:::
Db   179 SISVLISALIIYFKPEYKIADPICTFIFSILVLASTITILKDFSILLMEGVPKSLNYSGV 238

Qy   291 RDLLLSVEGVEALHSLHIWALTVAQPVLSVHIAIAQNTDAQAVLKTASSRLQGKFHFHTV 350
          :|:|||:|||||||:|||| | |||||||||||||||||||| |||:||||||||||||
Db   239 KELILAVDGVLSVHSLHIWSLTMNQVILSAHVATPAASRDSQVVRREIAKALSKSFTMHSL 298

Qy   351 TIQIEDYSEDMKDCQACQGPSD 372
          |||:| : |||||||||| |
Db   299 TIQMESPVDQDPDCLFCEDPCD 320
```

```
RESULT 8
AAE17562
ID   AAE17562 standard; Protein; 320 AA.
XX
AC   AAE17562;
XX
DT   22-APR-2002  (first entry)
XX
DE   Human pancreatic tumour full-length protein, 981418.1.
XX
KW   Human; pancreatic tumour protein; pancreatic cancer; gene therapy;
KW   vaccine.
XX
OS   Homo sapiens.
XX
PN   WO200194409-A2.
XX
PD   13-DEC-2001.
XX
PF   31-MAY-2001; 2001WO-US18003.
XX
PR   07-JUN-2000; 2000US-210329P.
PR   14-NOV-2000; 2000US-248980P.
PR   15-MAY-2001; 2001US-291197P.
XX
PA   (CORI-) CORIXA CORP.
XX
PI   Hirst SK,  Harlocker SL,  Dillon DC,  Kalos MD;
XX
DR   WPI; 2002-154565/20.
DR   N-PSDB; AAD28259.
XX
PT   Pancreatic tumor polypeptide and polynucleotide useful in diagnosis,
PT   prevention and/or treatment of cancer, especially pancreatic cancer
PS   Claim 2a; Page 111; 115pp; English.
```

```
CC   The invention relates to human pancreatic tumour polypeptides and
CC   nucleic acid molecules encoding such polypeptides. The invention
CC   also relates to compositions and methods for the diagnosis,
CC   prevention and therapy of cancer, particularly pancreatic cancer.
CC   Sequences of the invention are used in gene therapy and in vaccines.
CC   The present sequence is human pancreatic tumour full-length protein.
XX
SQ   Sequence   320 AA;

Query Match   46.9%;  Score 902.5; DB 23;  Length 320;
Best Local Similarity   53.1%;  Pred. No. 2.1e-88;
Matches  171;  Conservative   61; Mismatches   87;  Indels    3;  Gaps    1;

Qy    51 HHCHAQKGPDSHCDPKKGKAQRQLYVASAICLLFMIGEVVGGYLAHSLAVMTDAAHLLTD 110
          |:||:|                | |||     ||::|||:| |||||::||||||||||
Db     2 YHCHSGSKPTEKGANEYAYAKWKLCSASAICFIFMIAEVVGGHIAGSLAVVTDAAHLLID 61

Qy   111 FASMLISLFSLWMSSRPATKTMNFGWQRAEILGALVSVLSIWVVTGVLVYLAVERLISGD 170
          |||  |||||| ||:||  || |||||||||||:||| ||||||||| |  |:|||:||
Db    62 LTSFLLSLFSIWLSSKPPSKRLTFGWHRAEILGALLSILCIWVVTGVLVYLACERLLYPD 121

Qy   171 YEIDGGTMLITSGCAVANIIMGLTLHQSGHGHSHGTTNQQEENPSVRAAFIHVIGDFMQ 230
          |:|| | | ||| ::||  ||| |||| ||||||:|||     | :| |||||||||
Db   122 YQIQATVMIIVSSCAVANIVLTVVLHQRCLGHNH--KEVQANASVRAAFVHALGDLFQ 178

Qy   231 SMGVLVAAYILYFKPEYKYVDPICTFVFSILVLGTTLTILTLRDVILVLMEGTPKGVDFTAV 290
          :|||| ||| | |||:| |||||||||||||:|:|| | ||||||| :||| ||:::
Db   179 SISVLISALIIYFKPEYKIADPICTFIFSILVLASTITILKDFSILLMEGVPKSLNYSGV 238

Qy   291 RDLLLSVEGVEALHSLHIWALTVAQPVLSVHIAIAQNTDAQAVLKTASSRLQGKFHFHTV 350
          :|:|||:|||||||:|||| | |||||||||||||||||||| |||:||||||||||||
Db   239 KELILAVDGVLSVHSLHIWSLTMNQVILSAHVATPAASRDSQVVRREIAKALSKSFTMHSL 298

Qy   351 TIQIEDYSEDMKDCQACQGPSD 372
          |||:| : |||||||||| |
Db   299 TIQMESPVDQDPDCLFCEDPCD 320
```

```
RESULT 9
ABB60324
ID   ABB60324 standard; Protein; 472 AA.
XX
AC   ABB60324;
XX
DT   26-MAR-2002  (first entry)
XX
DE   Drosophila melanogaster polypeptide SEQ ID NO 7764.
XX
KW   Drosophila; developmental biology; cell signalling; insecticide;
KW   pharmaceutical.
XX
OS   Drosophila melanogaster.
XX
PN   WO200171042-A2.
XX
PD   27-SEP-2001.
XX
PF   23-MAR-2001; 2001WO-US09231.
XX
PR   23-MAR-2000; 2000US-191637P.
PR   11-JUL-2000; 2000US-0614150.
XX
PA   (PEKE ) PE CORP NY.
XX
PI   Venter JC,  Adams M,  Li PWD,  Myers EW;
XX
DR   WPI; 2001-656860/75.
DR   N-PSDB; ABL04427.
XX
PT   New isolated nucleic acid detection reagent for detecting 1000 or more
PT   genes from Drosophila and for elucidating cell signalling and cell-cell
PT   interactions -
```

```
XX
PS    Disclosure; SEQ ID NO 7764; 21pp + Sequence Listing; English.
XX
CC    The invention relates to an isolated nucleic acid detection reagent
CC    capable of detecting 1000 or more genes from Drosophila. The invention is
CC    useful in developmental biology and in elucidating cell signalling and
CC    cell-cell interactions in higher eukaryotes for the development of
CC    insecticides, therapeutics and pharmaceutical drugs. The invention
CC    discloses genomic DNA sequences (ABL01840-ABL06175), expressed DNA
CC    sequences (ABL01840-ABL06175) and the encoded proteins
CC    (ABB57737-ABB72072).
CC    The sequence data for this patent did not form part of the printed
CC    specification, but was obtained in electronic format directly from WIPO
CC    at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ    Sequence    472 AA;

Query Match          45.78%;  Score 878.5;  DB 22;  Length 472;
Best Local Similarity   42.6%;  Pred. No. 1.5e-85;
Matches  181;  Conservative   64;  Mismatches   83;  Indels   97;  Gaps    6;

Qy     36  PGLDLQAIELAAQSNH------HCH--AQKGPDSHCDPKKGKAQRQLYVASAICLLFMIGE  88
               ||| |:|||:|||||       |  |  | ||    |||||   |  |||:|  |
Db     53  PATPAQIFCLHGRSNNVEVRDHCHRARSEGVDV------KARRKLIIASICLVEMIAE  105

Qy     89  VVGGYLAHSLAVMTDAAHLLTDFASMLISLFSLWMSSRPAFKTMNFGWQRAEILGALVSV  148
               :||: ||||||||||: |||||||||||| |:||: ||| ||||:  :||| ||| |||
Db    106  IVGGVLSNSLAIATDAAHLLTDFASFMISLFAIWIAGRPSTQRMSFGWYRAEVIGAMASV  165

Qy    149  LSIWVVTGVLYVLYLAVERLLISGDYEIDGGTMLITSGCAVAVNIIMGLTLHQ------  201
               || |: | |||||||:|||  |:||:  |  |||||| |||||:||||||:|      |:|
Db    166  FMIWITGILVWLAIGRLLISGDYEVNAKIMLITSGLAILVNVIMGVQLQHGHSHGLGGGH  225

Qy    202  -------TTNQQEENPS------------  216
                       ||:|||||
Db    226  GHSHGGSKNASHVQATSTPCSDSPSQRIEGGVAYAPEDAELPGGGLPTFSYQNTKLVDPT  285

Qy    217  --------VRAAFIHVIGDFMQSMGVLVAAYILYFKP  245
                        |||:||||| | :|:|| |||||||:||
Db    286  LDLEIAAVLAETAPGSHHHGGPVGREAVNMNVRAALIHVIGDVIQSVGVFVAAGVIFWP  345

Qy    246  EYKYDVDPICFEVFSILVLGTTLILRDVILVLMEGTPKGVDFTAVRDLLSVEGVEALHS  305
               | |||:|| ||:||:|||:||||:||| || |||| | ||| :| :|:|:| ||| :|
Db    346  EYSIVDPICTFVFSIIVLFTFTIMKDALLVLMEGTPNYMHVAEVLCIFQGIEGVERVHN  405

Qy    306  LHIWALTVAQPVLSVHIAIAQNTDAQAVLKTASSRLQGKFHFHTVTIQIEDYSEDMKDCQ  365
               | |||| |||:|| |:|  |  |:: ||||  :||  |  |:  || :||:: :|||| |
Db    406  LRIWALSINKVLALSAHLAIAENANPKRILDAATSAVHLRYNFEFTTIQIEDYTAQMESCL  465

Qy    366  ACQGP  370
               | ||
Db    466  QCNVP  470
```

Human; NOVX; developmental disorder; endocrine disorder;
vascular disorder; infectious disease; anorexia; cancer; stroke;
neurodegenerative disorder; Alzheimer's disease; acute brain injury;
central nervous system disorder; depression; lung disorder;
reproductive disorder; tissue disorder; thrombocytopaenia; migraine;
angiogenesis; asthma; X-linked severe combined immunodeficiency;
inflammation; autoimmune disorders; immune disorder; blood disorder;
haematopoietic disorder; gastrointestinal disease; respiratory disorder;
hepatitis; fertility; hypertension; arteriosclerosis; ischaemia;

```
KW    rheumatoid arthritis; Grave's disease; wound healing.
XX
OS    Homo sapiens.
XX
PN    WO2002247333-A2.
XX
PD    28-MAR-2002.
XX
PF    17-SEP-2001;  2001WO-US29115.
XX
PR    15-SEP-2000;  2000US-232675P.
PR    15-SEP-2000;  2000US-232676P.
PR    15-SEP-2000;  2000US-232679P.
PR    18-SEP-2000;  2000US-233382P.
PR    18-SEP-2000;  2000US-234402P.
PR    19-SEP-2000;  2000US-233521P.
PR    19-SEP-2000;  2000US-233522P.
PR    19-SEP-2000;  2000US-233801P.
PR    20-SEP-2000;  2000US-233960P.
PR    06-OCT-2000;  2000US-238398P.
PR    06-OCT-2000;  2000US-238399P.
PR    13-OCT-2000;  2000US-240284P.
PR    13-OCT-2000;  2000US-240498P.
PR    11-JAN-2001;  2001US-260973P.
PR    26-JAN-2001;  2001US-264274P.
PR    09-MAR-2001;  2001US-274862P.
XX
PA    (CURA-) CURAGEN CORP.
XX
PI    Mishra VS, Syptek KA, Taupier RJ, Vernet CAM, Colman SD, Gorman L;
PI    Tcherney VT, Malyankar UM, Shenoy S, Tcherney VT, Padigaru M;
PI    Patturajan M, Burgess CE, Smithson G, Millet I, Peyman JA;
PI    Stone D, Gunther E, Ellerman K;
XX
DR    WPI; 2002-383182/41.
DR    N-PSDB; ABK71913.
XX
PT    New cytoplasmic, nuclear, membrane bound and secreted NOVX
PT    polypeptides, useful for treating cancers and tumours, lung disorders,
PT    haematopoietic disorders, autoimmune diseases and immune disorders -
XX
PS    Claim 1 ; Page 21; 210pp; English.
XX
CC    The invention relates to an isolated NOVX polypeptide selected from
CC    NOV1a, NOV1b, NOV1ac, NOV2a, NOV2b, NOV2c, NOV3a, NOV3b, NOV4a, NOV4b,
CC    NOV5a, NOV5b or NOV6-NOV9 polypeptides, their mature form or variant.
CC    Also included are a nucleic acid encoding a NOVX protein or variant;
CC    a vector comprising the nucleic acid; a cell comprising the vector;
CC    an anti-NOVX antibody; and identifying agents that modulate the
CC    expression or activity of NOVX. NOVX, the nucleic acid, antibody and
CC    modulators are useful in the diagnosis, treatment or prevention of
CC    developmental disorders, endocrine disorders, vascular disorders
CC    infectious disease, anorexia, cancer, neurodegenerative disorders (e.g.
CC    Alzheimer's disease, Parkinson's disease, Huntington's disease, multiple
CC    sclerosis and amyotropic lateral sclerosis), acute brain injury (e.g.
CC    stroke, head injury and cerebral palsy), central nervous system disorders
CC    (e.g. depression, epilepsy and schizophrenia), lung disorders and
CC    reproductive disorders, disorders affecting carbohydrate metabolism (e.g.
CC    galactosaemia and hereditary fructose intolerance), tissue disorders
CC    (e.g. Wiskott-aldrich syndrome, thrombocytopaenia, night blindness and
CC    Pick's disease), disorders linked to abnormal angiogenesis, asthma,
CC    azoospermia, learning disabilities, facial dysmorphism, autoimmune
CC    encephalomyelitis, X-linked severe combined immunodeficiency, seizures,
CC    migraines, inflammation, autoimmune disorders, disorders affecting sleep,
CC    appetite, thermoregulation, pain, perception, hormone secretion and
CC    sexual behaviour, immune disorders, haematopoietic disorders or other
CC    disorders related to cell signal processing and metabolic pathway
CC    modulation, gastrointestinal diseases, respiratory disorders,
CC    infections, hepatitis, trauma, regeneration, viral, bacterial or parasitic
CC    infections, hyper- or hypo-thyroidism, endometriosis, fertility,
CC    hypertension, arteriosclerosis, ischaemia, haemolytic anaemia, Werner
CC    syndrome, rheumatoid arthritis, Grave's disease, wound healing, X-linked
CC    mental retardation, psychotic and neurological disorders and neuronal
CC    degeneration. The present sequence represents a NOVX protein.
```

```
XX
SQ   Sequence    422 AA;
XX
Query Match     Best Local Similarity   43.6%;   Score 838.5;   DB 23;   Length 422;
Matches 166;   Conservative   47.8%;   Pred. No. 2.5e-81;   Mismatches 97;   Indels 15;   Gaps 4;

Qy     32  PLPRPGLDLQAIELAAQSNHHCHAQKGPDSHCDPKKGKAQRQLYVASAICLLFMIGEVVG  91
Db     38  PRERPE------ELESGGMYHCHSGSKPTEKGANEYAYAKWKLCSASAICFIFMIAEVVG  91

Qy     92  GYLAHSLAVMTDAAHLLTDFASMLISLFSLWMSSRPATKTMNFGWQRAEILGALVSVLSI  151
Db     92  GHIAGSLAAVVTDAAHLLIDLTFSFLLSLFSIWLSSKKPPSKRLTFGWHRAQVLFSILSLITL  151

Qy    152  WVVTGVLVYLAVERLLSGDYEIDGGTMLITSGCAVAV--NIIMG----LTLHOSGHGHSH  205
Db    152  VVVTGVLVYLACERLLYPDYQIQOATVMIIVSSCAVAAAKNIVLSFRLTVVLHQRCLGRNH  211

Qy    206  GTTNQGEENPSVRAAFIHVIGDFMQSMGVLVAAYILYFKPEYKYVDPICTFVFSILVLGT  265
Db    212  ---KEVOANASVRAAFVHALGDLFQSISVLISALIIYFKPEYKIADPICTFIFSILVLAS  268

Qy    266  TLTILRDVILVLMEGTPKGVDFTAVRDLLLSVEGVEALHSLHIWALTVAQPVLSVHIAIA  325
Db    269  TISILKDFFLLMEGVPKSLNYSGVKELILSVDGVLSVHSLHIWSLTMNQVILSAHVATA  328

Qy    326  QNTDAQAVLKTASSRLQGKFHFHTVTIQIEDYSEDMKDCQACQGPSD  372
Db    329  ASRDSQVVRREIAKALSKSFTMHSLTIQMESPVDQDPDCLFCEDPCE  375

RESULT 11
ABB69976
ID   ABB69976 standard; Protein; 1677 AA.
XX
AC   ABB69976;
XX
DT   26-MAR-2002  (first entry)
XX
DE   Drosophila melanogaster polypeptide SEQ ID NO 36720.
XX
KW   Drosophila; developmental biology; cell signalling; insecticide;
KW   pharmaceutical.
XX
OS   Drosophila melanogaster.
XX
PN   WO200171042-A2.
XX
PD   27-SEP-2001.
XX
PF   23-MAR-2001; 2001WO-US09231.
XX
PR   23-MAR-2000; 2000US-191637P.
PR   11-JUL-2000; 2000US-0614150.
XX
PA   (PEKE ) PE CORP NY.
XX
PI   Venter JC,  Adams M,  Li PWD,  Myers EW;
XX
DR   WPI; 2001-656860/75.
DR   N-PSDB; ABL14079.
XX
PT   New isolated nucleic acid detection reagent for detecting 1000 or more
PT   genes from Drosophila and for elucidating cell signalling and cell-cell
PT   interactions -
XX
PS   Disclosure; SEQ ID NO 36720; 21pp + Sequence Listing; English.
XX
CC   The invention relates to an isolated nucleic acid detection reagent
CC   capable of detecting 1000 or more genes from Drosophila. The invention is
CC   useful in developmental biology and in elucidating cell signalling and
CC   cell-cell interactions in higher eukaryotes for the development of
```

```
CC   insecticides, therapeutics and pharmaceutical drugs. The invention
CC   discloses genomic DNA sequences (ABL01840-ABL16175) and the encoded DNA
CC   sequences (ABL16176-ABL30511), expressed proteins
CC   (ABB57737-ABB72072).
CC   The sequence data for this patent did not form part of the printed
CC   specification, but was obtained in electronic format directly from WIPO
CC   at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ   Sequence    1677 AA;
XX
Query Match     Best Local Similarity   42.2%;   Score 811.5;   DB 22;   Length 1677;
Matches 172;   Conservative   37.3%;   Pred. No. 1.6e-77;   Mismatches 114;   Indels 95;   Gaps 6;

Qy      4  KEKOHLLDARPAIRSYT------------------------------------GSLWQEGAGW  30
Db    684  KNSENALDAKPQTATNTEETHGCINILKVPFQSNLIYYPEKPSRQDKGSGTSDFQPGAPE  743

Qy     31  IPLPRPGL---DLQAIELAAQS----------------------NHHCHAQKGPDSHCDPKK  67
Db    744  TPIVENSAVDSDRRAVEIMPENVKNSEEKKIDNSDSTKIVTITGHSHITAKWDGHCHRKE  803

Qy     68  GK------AQRQLYVASAICLLFMIGEVVGGYLAHSLAVMTDAAHLLTDFASMLISLFSL  121
Db    804  RETGVDKAAARRVLIIACILCTIFLILEVIGGILSNSLAIATDAAHLLTDLASFLISSAL  863

Qy    122  WMSSRPATKTMNFGWQRAEILGALVSVLSIWVVTGVLVYLAVERLISGDYEIDGGTMLIT  181
Db    864  HLAGRPSSERLNYGWHRAEVIGAMVSIFFIWVVTGILVYMAIMRWVNQDFELDAKIMLIT  923

Qy    182  SGCAVAVNIIMGLTLHQSGHGHSGTTN------------------------------QQ  211
Db    924  SALAILFNVIMAMOLQ---HGHSHSLPGVHKMSKDAGSVLGSKMILLGKSVSMOYAARG  980

Qy    212  EENPSVRAAFIHVIGDEMQSMGVLVAAYILYFKPEYKYVDPICTFVFSILVLGTTLTILR  271
Db    981  HENINVRAAIIHVVGDIIQOSFGVFVAALIIFFWPEWAFMDSVCTFVFSVLVLVVTFKILR  1040

Qy    272  DVILVLMEGTPKGVDFTAVRDLLLSVEGVEALHSLHIWALTVAQPVLSVHIAIAQNTDAQ  331
Db   1041  DVLMVLMEATPDFMDYEEVVKQTFLSISGVEHVHNLRIWALSINKVALSAHLAISKDADQ  1100

Qy    332  AVLKTASSRLQGKFHFHTVTIQIEDYSEDMKDCQACQGPSD  372
Db   1101  LILEEATLLIHKRFKFFETTIQIEEYSPGMENCQCLSPSD  1141

RESULT 12
ABG61811
ID   ABG61811 standard; Protein; 429 AA.
XX
AC   ABG61811;
XX
DT   15-AUG-2002  (first entry)
XX
DE   Prostate cancer-associated protein #12.
XX
KW   Prostate cancer; prostate tumour tissue; human; mammal; cytostatic.
XX
OS   Mammalia.
XX
PN   WO200230268-A2.
XX
PD   18-APR-2002.
XX
PF   12-OCT-2001; 2001WO-US32045.
XX
DT   15-AUG-2002  (first entry)
XX
PR   13-OCT-2000; 2000US-0687576.
PR   08-DEC-2000; 2000US-0733288.
PR   08-DEC-2000; 2000US-0733742.
PR   24-JAN-2001; 2001US-263957P.
PR   16-MAR-2001; 2001US-276791P.
PR   16-MAR-2001; 2001US-276888P.
```

```
PR   06-APR-2001; 2001US-2819221P.
PR   24-APR-2001; 2001US-286214P.
PR   30-APR-2001; 2001US-0847046.
PR   04-MAY-2001; 2001US-2885891P.
XX
PA   (EOSB-) EOS BIOTECHNOLOGY INC.
XX
PI   Gish KC, Mack DH, Wilson KE, Afar D, Hevezi P;
XX
DR   WPI; 2002-471335/50.
DR   N-PSDB; ABK92126.
XX
PT   Detecting a prostate cancer-associated transcript in a cell in a
PT   patient, useful for diagnosing prostate cancer (PC) or screening
PT   modulators of PC, by determining if prostate cancer-associated genes
PT   are expressed in a prostate tissue
XX
PS   Claim 27; Page 310; 436pp; English.
XX
CC   The present invention relates to methods of detecting a prostate
CC   cancer-associated transcript in a cell from a patient. The method
CC   comprises contacting a biological sample from the patient with
CC   prostate cancer-associated polynucleotides (designated PC genes) that
CC   selectively hybridise to a sequence that is at least 80% identical
CC   to them. The prostate cancer-associated polynucleotide sequences
CC   are differentially expressed in prostate tumour tissue or in
CC   prostate cancer and are derived from the tissues of various
CC   organisms such as humans or other mammals (e.g. mice, sheep and dogs).
CC   The methods of the invention are useful for diagnosing and treating
CC   prostate cancer in mammals. The prostate cancer-associated genes are
CC   useful for diagnosing or treating prostate cancer, as well as for
CC   identifying modulators of prostate cancer or agents that inhibit
CC   prostate cancer. The nucleic acid sequences are particularly useful
CC   in gene therapy, as a vaccine or in antisense applications.
     ABG61800-ABG61944 represent prostate cancer-associated proteins.
XX
SQ   Sequence        429 AA;

Query Match             37.6%;  Score 722.5;  DB 23;  Length 429;
Best Local Similarity   39.0%;  Pred. No. 8.1e-69;
Matches  151;  Conservative   88;  Mismatches 117;  Indels   31;  Gaps  7;

Qy      2 EAKEKQHLLDARPAIRSYTGSLWQEGAGWIPLPRPGLDLQAIELAAQSNHHCHAQKGPDS 61
          ||  |  :   | |  :: | ||    ::      | | | ::    :  |  | ||
Db     56 EAPERP-VNGAHPTLQADDDSLLDQD---LPLTNSQLSLK-----VDSCDNCSKQRE--- 103

Qy     62 HCDPKKGKAQRQLYVASAICLLFMIGEVVGGYLAHSLAVWTDAAHLLTDFASMLISLFSL 121
             |: | | :|| :|| |||||||||:|||||:| ||| ::|||  |||| :: | |  |
Db    104 --ILKQRKVKARLTIAAVLYLLFMIGELVGGYIANSLAIMTDALHMLTDLSAIILTLLAL 161

Qy    122 WMSSRPATKTMNFGWQRAEILGALVSVLSIWVVTGVLVLVYLAVERLISGDYEIDGGTMLIT 181
          | ||  | || | | :|| :|| :|| ||:  : |||:| | ||  | ::| ::  || ||||
Db    162 WLSSKSPTKRFTFGFHRLEVLSAMISVLLVYIILMGFLLYEAVQRTIHMNYEINGDIMLIT 221

Qy    182 SGCAVAVNIIMGLTLHQSGHGHSHGTT----------- NQQEENPSVRAAFIHVIGD 227
           | ||||| ||| |:|:|||:|||||              ::| |||||||| || ||
Db    222 AAVGVAVNVIMGFLLNQSGRHSHSHSLPSNSPTRGSGCERNHGQDSLAVRAAFVHALGD 281

Qy    228 FMQSMGVLVAAVILYFKPEYKVDPICTFVFSILVLGTTLTILRDVILVLMEGTPKGVDF 287
           :|| | |:|| :|| |||||:|||||:|||:||| ||| |:|||| ::|:|| | || |
Db    282 LVQSGVGLIAAYIIRFKPEYKIADPICTYVFSLLVAFTTFRIIWDTVVIILEGVPSHLNV 341

Qy    288 TAVRDLLLSVEGVEALHSLHIWALTVAQPVLSVHIAIAQNTDA--QAVLKTASSRLQGKF 345
           :: | :| :|| || || ||:||| || : |:|||: |   |    || | |:| | |
Db    342 DYIKEALMKIEDVYSVEDLNIWSLISGKSTAIVHIQLIPGSSSKWEEVQSKANHLLLNTF 401

Qy    346 HFHTVHIQIEDYSEDM-KDCQACGPS 371
          :  :|  : |:| |||   :| :| |
Db    402 GMYRCTIQLQSYRQEVDRTCANCQSSS 428
```

```
ID   AAE24062 standard; Protein; 438 AA.
XX
AC   AAE24062;
XX
DT   23-SEP-2002 (first entry)
XX
DE   Human prostate specific protein (PSP) #5.
XX
KW   Human; prostate specific protein; PSP; prostate specific nucleic acid;
KW   vaccine; transgenic; prostate cancer; gene therapy; transgenic animal;
KW   PSNA.
XX
OS   Homo sapiens.
XX
XX
PN   WO20224718-A1.
XX
PI   Sun Y, Recipon H, Cafferkey R, Ali S;
XX
XX
PR   19-SEP-2001; 2001WO-US29386.
PR   19-SEP-2000; 2000US-233746P.
XX
PA   (DIAD-) DIADEXUS INC.
XX
PT   Novel isolated prostate specific polypeptide useful for identifying,
PT   diagnosing, monitoring, staging, imaging, and treating prostate cancer
PT   and non-cancerous disease states in prostate -
XX
PS   Claim 37; Page 198-199; 210pp; English.
XX
CC   The invention relates to prostate specific proteins (PSP) and prostate
CC   specific nucleic acids (PSNA). Sequences of the invention are useful
CC   for identifying, diagnosing, monitoring, staging, imaging and treating
CC   prostate cancer and non-cancerous disease states in prostate. They are
CC   also useful for producing engineered prostate tissue for treatment and
CC   research. The PSNA sequences are used in gene therapy and for producing
CC   transgenic animals and cells. The invention is also used as vaccines.
CC   The present sequence is human prostate specific protein of the invention.
XX
SQ   Sequence        438 AA;

Query Match             37.6%;  Score 722.5;  DB 23;  Length 438;
Best Local Similarity   39.0%;  Pred. No. 8.4e-69;
Matches  151;  Conservative   88;  Mismatches 117;  Indels   31;  Gaps  7;

Qy      2 EAKEKQHLLDARPAIRSYTGSLWQEGAGWIPLPRPGLDLQAIELAAQSNHHCHAQKGPDS 61
          ||  |  :   | |  :: | ||    ::      | | | ::    :  |  | ||
Db     56 EAPERP-VNGAHPTLQADDDSLLDQD---LPLTNSQLSLK-----VDSCDNCSKQRE--- 103

Qy     62 HCDPKKGKAQRQLYVASAICLLFMIGEVVGGYLAHSLAVWTDAAHLLTDFASMLISLFSL 121
             |: | | :|| :|| |||||||||:|||||:| ||| ::|||  |||| :: | |  |
Db    104 --ILKQRKVKARLTIAAVLYLLFMIGELVGGYIANSLAIMTDALHMLTDLSAIILTLLAL 161

Qy    122 WMSSRPATKTMNFGWQRAEILGALVSVLSIWVVTGVLVLVYLAVERLISGDYEIDGGTMLIT 181
          | ||  | || | | :|| :|| :|| ||:  : |||:| | ||  | ::| ::  || ||||
Db    162 WLSSKSPTKRFTFGFHRLEVLSAMISVLLVYIILMGFLLYEAVQRTIHMNYEINGDIMLIT 221

Qy    182 SGCAVAVNIIMGLTLHQSGHGHSHGTT----------- NQQEENPSVRAAFIHVIGD 227
           | ||||| ||| |:|:|||:|||||              ::| |||||||| || ||
Db    222 AAVGVAVNVIMGFLLNQSGRHSHSHSLPSNSPTRGSGCERNHGQDSLAVRAAFVHALGD 281

Qy    228 FMQSMGVLVAAVILYFKPEYKVDPICTFVFSILVLGTTLTILRDVILVLMEGTPKGVDF 287
           :|| | |:|| :|| |||||:|||||:|||:||| ||| |:|||| ::|:|| | || |
Db    282 LVQSGVGLIAAYIIRFKPEYKIADPICTYVFSLLVAFTTFRIIWDTVVIILEGVPSHLNV 341

Qy    288 TAVRDLLLSVEGVEALHSLHIWALTVAQPVLSVHIAIAQNTDA--QAVLKTASSRLQGKF 345
           :: | :| :|| || || ||:||| || : |:|||: |   |    || | |:| | |
Db    342 DYIKEALMKIEDVYSVEDLNIWSLTSGKSTAIVHIQLIPGSSSKWEEVQSKANHLLLNTF 401
```

```
RESULT 13
AAE24062
```

Left column

```
Qy   346  HFHTVTIQIEDYSEDM-KDCQACQGPS 371
          ||::|| ||:::| ::| | :  ||| ||
Db   402  GMYRCTIQLQSYRQEVDRTCANCQSSS 428
```

```
RESULT 14
AAE09322
ID   AAE09322 standard; Protein; 474 AA.
XX
AC   AAE09322;
XX
DT   22-NOV-2001  (first entry)
XX
DE   Rice zinc transporter (ZnT-2) protein from clone r10n.pk0012.c11:fis.
XX
KW   Heavy metal transporter; zinc transporter; transgenic plant;
KW   rice; ZnT-2.
XX
OS   Oryza sativa.
XX
XX
XX
PN   US6278042-B1.
XX
PD   21-AUG-2001.
XX
PF   14-DEC-1999;   99US-0461474.
XX
PR   16-DEC-1998;   98US-0112562.
XX
PA   (DUPO ) DU PONT DE NEMOURS & CO E I.
XX
PI   Allen SM, Rafalski JA, Sakai H;
XX
DR   WPI; 2001-540412/60.
DR   N-PSDB; AAD16066.
XX
PT   New arsenic transporter polypeptides and polynucleotides, useful for
PT   producing transgenic plants with altered level of heavy metal tolerance
-
XX
CC   The invention relates to heavy metal (e.g. arsenite and zinc)
CC   transporter polypeptides and polynucleotides. The polypeptides of
CC   the invention are useful for producing antibodies that are useful
CC   for screening cDNA expression libraries to isolate full length
CC   clones of interest. The antibodies are also useful for detecting
CC   the polypeptides in situ, in cells or in vitro, in cell extracts.
CC   Nucleic acid molecules of the invention are useful for producing
CC   transgenic plants with altered tolerance to heavy metals. They are
CC   also useful as probes for genetically and physically mapping the
CC   genes that are a part of, and as markers for traits linked to those
CC   genes. The present sequence is rice zinc transporter (ZnT-2) protein.
XX
PS   Example 4; Fig 2; 31pp; English.
XX
SQ   Sequence   474 AA;
```

```
Query Match       34.1%;  Score 656.5;  DB 22;  Length 474;
Best Local Similarity   31.6%;   Pred. No. 1.2e-61;
Matches  151;  Conservative  72;  Mismatches  98;  Indels  157;  Gaps  8;

Qy    17  RSYTGSLWQEGAGWIPLPRPGLDLQAIELAAQSNHHCHAQKGPDSHCD------------  64
              :||:||  ::          ||||||
Db    13  RAYTRSL------LPISN------ARTRHHHLDAGGDDHGDNGGGGREALLI  54

Qy    65  PK----------------------------KGKA------------QRQL  74
              ||                          ||
Db    55  PKMDSHNSAPPQIAEVRMDISSSTSVAAGNKVCRGAACDFSDSSNSSKDARERMASMRKL  114

Qy    75  YVASAICLLFMIGEVVGGYLAHSLAVWTDAAHLLTDFASMLISLFSLWMSSRPATKTMNF  134
              :||:  :|| | ||||||   ||:||||||||||||||||||||:   ::
Db   115  IIAVILCIIFMAVEVGGGIKANSLAILTDAAHLLSDVAAFAISLFSLWAAGWEATPQQSY  174

Qy   135  GWQRAEILGALVSVLSIWVVTGVLVILAVERLISGDYEIDGGTMLITSGCAVAVNIIMGL  194
          |:| | |||||||: |:|||||||| ||:|| |||||:|||:||| | ||||:|  :
```

Right column

```
Db   175  GFPRIEILGALVSIQLIWLLAGIVYEAIVRLINESGEVQGSLMFAVSAFGLFVNIIMAV  234
Qy   195  TL-HQSGHGSHG--------------------------------------TTN----  209
          || | | |||| ||                                    || |
Db   235  LLGHDHGHGHGHGHSHDHDRGGSDHDHHHHEDQEHGHVHHHEDGHGNSITVNLHHH  294

Qy   210  --------------------------QQEENPSVRAAFIHVIGDFMQSMGVLVA  237
                                     | :| |:|:||||::|||:| ::|: ::
Db   295  PGTGHHHHDDAEEPLLKSDAGCDSTQSGAKDAKKARRNINVHSAYLHVLGDSIQSIGVMIG  354

Qy   238  AYILYFKPEYKYVDPICTFVFSILVLGTTGTTLILRDVILVLMEGTPKGVDFTAVRDLILSV  297
          | ||:||||:| ||| || |||:|||| ||  |:||:|:|||| || ||  |  : |
Db   355  GAIIWYKPEWKIIDLICTLIFSVIVLFTTIKMLRNILEVLMESTPREIDATISLENGLRDM  414

Qy   298  EGVEALHSLHIWALTVAQPVLSVHIAIAQNTDAQAVLKTPASSRLQGKFHFHTVTIQIE  355
          :|| |:| ||||:||  :||| |  |: | || |  ||:    :   |:|| |||||
Db   415  DGVVAVHELHIWAITVGKVLLACHVTITQDADADQMLDKVIGIYIKSEYNISHVTIQIE  472
```

```
RESULT 15
AAE09321
ID   AAE09321 standard; Protein; 448 AA.
XX
AC   AAE09321;
XX
DT   22-NOV-2001  (first entry)
XX
DE   Maize zinc transporter (ZnT-2) protein from clone cdt2c.pk002.h12:fis.
XX
KW   Heavy metal transporter; zinc transporter; transgenic plant;
KW   maize; ZnT-2.
XX
OS   Zea mays.
XX
XX
XX
PN   US6278042-B1.
XX
PD   21-AUG-2001.
XX
PF   14-DEC-1999;   99US-0461474.
XX
PR   16-DEC-1998;   98US-0112562.
XX
PA   (DUPO ) DU PONT DE NEMOURS & CO E I.
XX
PI   Allen SM, Rafalski JA, Sakai H;
XX
DR   WPI; 2001-540412/60.
DR   N-PSDB; AAD16065.
XX
PT   New arsenic transporter polypeptides and polynucleotides, useful for
PT   producing transgenic plants with altered level of heavy metal tolerance
-
XX
CC   The invention relates to heavy metal (e.g. arsenite and zinc)
CC   transporter polypeptides and polynucleotides. The polypeptides of
CC   the invention are useful for producing antibodies that are useful
CC   for screening cDNA expression libraries to isolate full length
CC   clones of interest. The antibodies are also useful for detecting
CC   the polypeptides in situ, in cells or in vitro, in cell extracts.
CC   Nucleic acid molecules of the invention are useful for producing
CC   transgenic plants with altered tolerance to heavy metals. They are
CC   also useful as probes for genetically and physically mapping the
CC   genes that are a part of, and as markers for traits linked to those
CC   genes. The present sequence is maize zinc transporter (ZnT-2) protein.
XX
PS   Example 4; Fig 2; 31pp; English.
XX
SQ   Sequence   448 AA;
```

```
Query Match       33.9%;  Score 651.5;  DB 22;  Length 448;
Best Local Similarity   35.3%;   Pred. No. 3.9e-61;
Matches  142;  Conservative  71;  Mismatches  100;  Indels  89;  Gaps  6;

Qy    17  RSYTGSLWQEGAGWIPLPRPGLDLQAIELAAQSNHHCHAQKGPDSHCD-----------  64
              :|:
Db    13  RAYTRSL-------------ARTRHHHLDAGGDDHGDNGGGGREALLI  54

Qy    65  PK----------------------------KGKA------------QRQL  74
              ||                          ||
Db    55  PKMDSHNSAPPQIAEVRMDISSSTSVAAGNKVCRGAACDFSDSSNSSKDARERMASMRKL  114

Qy    75  YVASAICLLFMIGEVVGGYLAHSLAVWTDAAHLLTDFASMLISLFSLWMSSRPATKTMNF  134
              :||:  :|| | ||||||   ||:||||||||||||||||||||:   ::
Db   115  IIAVILCIIFMAVEVGGGIKANSLAILTDAAHLLSDVAAFAISLFSLWAAGWEATPQQSY  174

Qy   135  GWQRAEILGALVSVLSIWVVTGVLVILAVERLISGDYEIDGGTMLITSGCAVAVNIIMGL  194
          |:| | |||||||: |:|||||||| ||:|| |||||:|||:||| | ||||:|  :
```

```
Qy    38  LDLQAIELAAQSNHHCHAQKGPDSHCD----------PKKGKAQRQLYVASAICLLFMI   86
          :||: ||  || ||    |      |   :| | :||:||
Db    50  MDIAASASGAAGSKPC--KG--AACDFSDASNSSKDARERSASMRKLIVAVVLCVVFMA  104

Qy    87  GEVGGYLAHSLAVMTDAAHLLTDFASMLISLFSLWMSSRPATKTMNFGWQRAEILGALV  146
          ||||| | |   || |||||| | |||:||||||  || :| | |||||||:||||:||
Db   105  VEVVGGIKANSLAILTDAAHLLSDVAAFAISLFSLWAAGWEATPRQSYGFFRVEILGALV  164

Qy   147  SVLSTWVTGVLVLYLAVERLISGDYEIDGGTMLITSGCAVAVNIIMGLTL--HQSGHGHSH  205
          | ||  |||:|:| |||| | |  ||||  |:|:|  | ||||:| |    |:||||||
Db   165  SIQLIWLLAGTLVFEAVVRLVGESGDVRGSLMFAVSAFGLAVNVLMAVLGHDHGHGHGH  224

Qy   206  G---------TTNQQEE-------------------  213
          :||:
Db   225  GHGHSHDHGHGHGDSDDGHSHHDDEEQEQGRVHHHEHGHGGAITVTTHHHHHHHDHDVEE  284

Qy   214  ------------NPSVRAAFIHVIGDFMQSMGVLVAAYILYFKPEYKYVDPI  253
              :| :||: :| :|:: | :| |:|
Db   285  ALIKHEGTQSAGRAGKKPRRNINVHSAYLHVLGDSVQSVGVMVGGAIIWKPEWKVIDLI  344

Qy   254  CTFVFSILVLGTTFLTILRDVILVLMEGTPKGVDFTAVRDLLLSVEGVEALHSLHIWALTV  313
          |||:||| || ::|| :| | ||:|:| || |:|  :|:|||:||::|:|||||||:
Db   345  CTLVPSVVLPTTIRMLRSILEVLMESTPREIDATRLESGLCGMEGVVAVHELHIWAITV  404

Qy   314  AQPVLSVHIAIAQNTDAQAVLKTASSRLQGKFHFHTVTIQIE  355
          |||| | :|:| |  : :|  |||| |||||:| :||:|:
Db   405  GKVLLACHVTIARDADADEILDKVIGYIKTEYNISHVTIQVE  446
```

RESULT 16

AAG22263

ID   AAG22263 standard; Protein; 398 AA.
XX
AC   AAG22263;
XX
DT   17-OCT-2000 (first entry)
XX
DE   Arabidopsis thaliana protein fragment SEQ ID NO: 25124.
XX
KW   Protein identification; signal transduction pathway; metabolic pathway;
KW   hybridisation assay; genetic mapping; gene expression control; promoter;
KW   termination sequence.
XX
OS   Arabidopsis thaliana.
XX
PN   EP1033405-A2.
XX
PD   06-SEP-2000.
XX
PF   25-FEB-2000; 2000EP-0301439.

PR   25-FEB-1999; 99US-0121825.
PR   05-MAR-1999; 99US-0123180.
PR   09-MAR-1999; 99US-0123548.
PR   21-APR-1999; 99US-0125788.
PR   23-MAR-1999; 99US-0126264.
PR   25-MAR-1999; 99US-0126785.
PR   29-MAR-1999; 99US-0126785.
PR   01-APR-1999; 99US-0127462.
PR   06-APR-1999; 99US-0128224.
PR   08-APR-1999; 99US-0128714.
PR   16-APR-1999; 99US-0129845.
PR   19-APR-1999; 99US-0130077.
PR   21-APR-1999; 99US-0130449.
PR   23-APR-1999; 99US-0130510.
PR   28-APR-1999; 99US-0130891.
PR   30-APR-1999; 99US-0131449.
PR   04-MAY-1999; 99US-0132048.
PR   05-MAY-1999; 99US-0132407.
PR   06-MAY-1999; 99US-0132484.
PR   06-MAY-1999; 99US-0132485.
PR   07-MAY-1999; 99US-0132486.
PR   07-MAY-1999; 99US-0132487.
PR   07-MAY-1999; 99US-0132863.

PR   11-MAY-1999; 99US-0134256.
PR   14-MAY-1999; 99US-0134218.
PR   14-MAY-1999; 99US-0134219.
PR   14-MAY-1999; 99US-0134221.
PR   14-MAY-1999; 99US-0134370.
PR   18-MAY-1999; 99US-0134768.
PR   19-MAY-1999; 99US-0134941.
PR   20-MAY-1999; 99US-0135124.
PR   21-MAY-1999; 99US-0135353.
PR   24-MAY-1999; 99US-0135629.
PR   25-MAY-1999; 99US-0136021.
PR   27-MAY-1999; 99US-0136392.
PR   28-MAY-1999; 99US-0136782.
PR   01-JUN-1999; 99US-0137222.
PR   03-JUN-1999; 99US-0137528.
PR   04-JUN-1999; 99US-0137502.
PR   07-JUN-1999; 99US-0137724.
PR   08-JUN-1999; 99US-0138094.
PR   10-JUN-1999; 99US-0138540.
PR   10-JUN-1999; 99US-0138847.
PR   14-JUN-1999; 99US-0139119.
PR   16-JUN-1999; 99US-0139110.
PR   18-JUN-1999; 99US-0139817.
PR   18-JUN-1999; 99US-0139899.
PR   21-JUN-1999; 99US-0139750.
PR   18-JUN-1999; 99US-0139463.
PR   18-JUN-1999; 99US-0139462.
PR   18-JUN-1999; 99US-0139461.
PR   18-JUN-1999; 99US-0139460.
PR   18-JUN-1999; 99US-0139459.
PR   18-JUN-1999; 99US-0139458.
PR   18-JUN-1999; 99US-0139457.
PR   18-JUN-1999; 99US-0139456.
PR   18-JUN-1999; 99US-0139455.
PR   18-JUN-1999; 99US-0139454.
PR   17-JUN-1999; 99US-0139492.
PR   16-JUN-1999; 99US-0139453.
PR   16-JUN-1999; 99US-0139452.
PR   14-JUN-1999; 99US-0139817.
PR   10-JUN-1999; 99US-0138847.
PR   08-JUN-1999; 99US-0138094.
PR   07-JUN-1999; 99US-0137724.
PR   04-JUN-1999; 99US-0137502.
PR   03-JUN-1999; 99US-0137528.
PR   01-JUN-1999; 99US-0137222.
PR   28-MAY-1999; 99US-0136782.
PR   27-MAY-1999; 99US-0136392.
PR   25-MAY-1999; 99US-0136021.
PR   24-MAY-1999; 99US-0135629.
PR   21-MAY-1999; 99US-0135353.
PR   20-MAY-1999; 99US-0135124.
PR   19-MAY-1999; 99US-0134941.
PR   18-MAY-1999; 99US-0134768.
PR   14-MAY-1999; 99US-0134370.
PR   14-MAY-1999; 99US-0134221.
PR   14-MAY-1999; 99US-0134219.
PR   14-MAY-1999; 99US-0134218.
PR   11-MAY-1999; 99US-0134256.
PR   09-JUL-1999; 99US-0142920.
PR   12-JUL-1999; 99US-0142977.
PR   13-JUL-1999; 99US-0143542.
PR   14-JUL-1999; 99US-0143624.
PR   15-JUL-1999; 99US-0144005.
PR   16-JUL-1999; 99US-0144085.
PR   16-JUL-1999; 99US-0144086.
PR   16-JUL-1999; 99US-0144325.
PR   19-JUL-1999; 99US-0144331.
PR   19-JUL-1999; 99US-0144332.
PR   19-JUL-1999; 99US-0144333.
PR   19-JUL-1999; 99US-0144334.
PR   19-JUL-1999; 99US-0144335.
PR   20-JUL-1999; 99US-0144632.
PR   20-JUL-1999; 99US-0144814.
PR   21-JUL-1999; 99US-0144884.
PR   21-JUL-1999; 99US-0145086.
PR   21-JUL-1999; 99US-0145088.
PR   22-JUL-1999; 99US-0145085.
PR   22-JUL-1999; 99US-0145087.
PR   22-JUL-1999; 99US-0145089.
PR   22-JUL-1999; 99US-0145192.
PR   23-JUL-1999; 99US-0145145.

```
PR   23-JUL-1999;   99US-0145218.
PR   23-JUL-1999;   99US-0145224.
PR   26-JUL-1999;   99US-0145276.
PR   27-JUL-1999;   99US-0145913.
PR   27-JUL-1999;   99US-0145918.
PR   28-JUL-1999;   99US-0145919.
PR   27-JUL-1999;   99US-0145951.
PR   02-AUG-1999;   99US-0146386.
PR   02-AUG-1999;   99US-0146388.
PR   02-AUG-1999;   99US-0146389.
PR   03-AUG-1999;   99US-0147038.
PR   04-AUG-1999;   99US-0147204.
PR   04-AUG-1999;   99US-0147302.
PR   05-AUG-1999;   99US-0147192.
PR   05-AUG-1999;   99US-0147260.
PR   06-AUG-1999;   99US-0147303.
PR   06-AUG-1999;   99US-0147416.
PR   09-AUG-1999;   99US-0147493.
PR   09-AUG-1999;   99US-0147935.
PR   10-AUG-1999;   99US-0148171.
PR   11-AUG-1999;   99US-0148319.
PR   12-AUG-1999;   99US-0148341.
PR   13-AUG-1999;   99US-0148565.
PR   13-AUG-1999;   99US-0148684.
PR   16-AUG-1999;   99US-0149368.
PR   17-AUG-1999;   99US-0149175.
PR   18-AUG-1999;   99US-0149426.
PR   20-AUG-1999;   99US-0149722.
PR   20-AUG-1999;   99US-0149723.
PR   20-AUG-1999;   99US-0149929.
PR   23-AUG-1999;   99US-0149902.
PR   23-AUG-1999;   99US-0149930.
PR   25-AUG-1999;   99US-0150566.
PR   26-AUG-1999;   99US-0150884.
PR   27-AUG-1999;   99US-0151065.
PR   27-AUG-1999;   99US-0151066.
PR   27-AUG-1999;   99US-0151080.
PR   30-AUG-1999;   99US-0151303.
PR   31-AUG-1999;   99US-0151438.
PR   01-SEP-1999;   99US-0151930.
PR   07-SEP-1999;   99US-0152363.
PR   10-SEP-1999;   99US-0153070.
PR   13-SEP-1999;   99US-0153758.
PR   15-SEP-1999;   99US-0154018.
PR   16-SEP-1999;   99US-0154039.
PR   20-SEP-1999;   99US-0154779.
PR   22-SEP-1999;   99US-0155139.
PR   23-SEP-1999;   99US-0155486.
PR   24-SEP-1999;   99US-0155659.
PR   28-SEP-1999;   99US-0156458.
PR   29-SEP-1999;   99US-0156596.
PR   04-OCT-1999;   99US-0157117.
PR   05-OCT-1999;   99US-0157753.
PR   06-OCT-1999;   99US-0157865.
PR   07-OCT-1999;   99US-0158029.
PR   08-OCT-1999;   99US-0158232.
PR   12-OCT-1999;   99US-0158369.
PR   13-OCT-1999;   99US-0159293.
PR   13-OCT-1999;   99US-0159294.
PR   13-OCT-1999;   99US-0159295.
PR   14-OCT-1999;   99US-0159329.
PR   14-OCT-1999;   99US-0159330.
PR   14-OCT-1999;   99US-0159331.
PR   14-OCT-1999;   99US-0159584.
PR   18-OCT-1999;   99US-0159637.
PR   18-OCT-1999;   99US-0159638.
PR   09-MAR-1999;   99US-0160741.
PR   09-MAR-1999;   99US-0160767.
PR   21-OCT-1999;   99US-0160768.
PR   21-OCT-1999;   99US-0160770.
PR   21-OCT-1999;   99US-0160814.
PR   21-OCT-1999;   99US-0160815.
PR   22-OCT-1999;   99US-0160980.
```

```
PR   22-OCT-1999;   99US-0160981.
PR   22-OCT-1999;   99US-0160989.
PR   25-OCT-1999;   99US-0161404.
PR   25-OCT-1999;   99US-0161405.
PR   25-OCT-1999;   99US-0161406.
PR   26-OCT-1999;   99US-0161359.
PR   26-OCT-1999;   99US-0161360.
PR   26-OCT-1999;   99US-0161361.
PR   28-OCT-1999;   99US-0161920.
PR   28-OCT-1999;   99US-0161992.
PR   28-OCT-1999;   99US-0161993.
PR   29-OCT-1999;   99US-0162142.
```

```
RESULT 17
AAG43478
ID   AAG43478 standard; Protein; 398 AA.
XX
AC   AAG43478;
XX
DT   18-OCT-2000  (first entry)
XX
DE   Arabidopsis thaliana protein fragment SEQ ID NO: 54349.
XX
KW   Protein identification; signal transduction pathway; metabolic pathway;
KW   hybridisation assay; genetic mapping; gene expression control; promoter;
KW   termination sequence.
XX
OS   Arabidopsis thaliana.
XX
XX
PN   EP1033405-A2.
XX
PD   06-SEP-2000.
XX
PF   25-FEB-2000; 2000EP-0301439.
XX
PR   25-FEB-1999;   99US-0218825.
PR   05-MAR-1999;   99US-0123180.
PR   09-MAR-1999;   99US-0123548.
PR   23-MAR-1999;   99US-0125788.
PR   25-MAR-1999;   99US-0126264.
PR   29-MAR-1999;   99US-0127785.
PR   01-APR-1999;   99US-0127462.
PR   06-APR-1999;   99US-0128234.
PR   08-APR-1999;   99US-0128714.
```

```
Query Match             33.8%;  Score  649.5;  DB 21;  Length 398;
Best Local Similarity   37.9%;  Pred. No. 5.3e-61;
Matches  134;  Conservative   71;  Mismatches   88;  Indels   61;  Gaps   4;

Qy   60 DSHCDPKKGKAQRQLYVASAICLLFMIGEVVGGYLAHSLAVMTDAAHLLTDFASHLISLF 119
            |:|  :   |:|  : :  | |||   :::||::::|||:|||::  |||  |||
Db   47 DAH--ERSASMRKLCIAVVLCLVFMSVEVVGGIKANSLAILTDAAHLLSDVAAFAISLF 103

Qy  120 SLWMSSRPATKTMNFGWQRAEILGALVSVLSIWVVTGVLVYLAVERLLISGDYEIDGGTML 179
           ||||  |  |||| ::| | || | |  ||||| |:|| |:||| |::| :  |||
Db  104 SLWAAGWEATPRQTGFFRIEILGALVSIQLIWLLTGILVYEAIIRIVTETSEVNGFLMF 163

Qy  180 ITSGCAVAVNIIMGLTL------HQSGHGHSHG----------- 209
          ::  | |||| ||::|                  TTN
Db  164 LVAAFGLVVNIIMAVLLGHDHGHSHGHGHDHHNHSHGVTVTTHHHHDHEHGHSHG 223

Qy  210 ----------------QOEENPSVRAAFIHVIGDFMQSMGVLVAAYIL 241
                          : |::   :||:|: ::||:||::
Db  224 HGEDKHHAHGDVTEQLLDKSKTQVAAKEKRKRNINLQGAYLHVLGDSIQSVGVMIGGAII 283

Qy  242 YFKPEYKYDPICIFEVFSILVLGTTLTILRDVILVLMEGTPKGVDFTAVRDLLLSVEGVE 301
          :: ||:|| ||  ||| |:|||:::|| |||| |:|: :|  |
Db  284 WYNPEWKIVDLICTLAFSVIVLGTTINMIRNILEVLMESTPREIDATKLEKGLLEMEEVV 343

Qy  302 ALHSLHIWALTVAQPVLSVHIAIAQNTDAQAVLKTASSRLQGKFHFHTVTIQIE 355
          |||||||||:|  : :|| |::   ::   | :||
Db  344 AVHELHIWAITVGKVLLACHVNIRPEADADMVLNKVIDYIRREYNISHVTIQIE 397
```

```
PR  16-APR-1999;  99US-0129845.
PR  19-APR-1999;  99US-0130077.
PR  21-APR-1999;  99US-0130449.
PR  23-APR-1999;  99US-0130510.
PR  28-APR-1999;  99US-0130891.
PR  28-APR-1999;  99US-0131449.
PR  30-APR-1999;  99US-0132048.
PR  30-APR-1999;  99US-0132407.
PR  04-MAY-1999;  99US-0132484.
PR  05-MAY-1999;  99US-0132485.
PR  06-MAY-1999;  99US-0132486.
PR  06-MAY-1999;  99US-0132487.
PR  07-MAY-1999;  99US-0132863.
PR  11-MAY-1999;  99US-0134256.
PR  14-MAY-1999;  99US-0134218.
PR  14-MAY-1999;  99US-0134219.
PR  14-MAY-1999;  99US-0134221.
PR  14-MAY-1999;  99US-0134370.
PR  18-MAY-1999;  99US-0134768.
PR  19-MAY-1999;  99US-0134941.
PR  20-MAY-1999;  99US-0135124.
PR  21-MAY-1999;  99US-0135353.
PR  24-MAY-1999;  99US-0135629.
PR  25-MAY-1999;  99US-0136021.
PR  27-MAY-1999;  99US-0136392.
PR  28-MAY-1999;  99US-0136782.
PR  01-JUN-1999;  99US-0137222.
PR  03-JUN-1999;  99US-0137528.
PR  04-JUN-1999;  99US-0137502.
PR  07-JUN-1999;  99US-0137724.
PR  08-JUN-1999;  99US-0138094.
PR  10-JUN-1999;  99US-0138540.
PR  10-JUN-1999;  99US-0138847.
PR  14-JUN-1999;  99US-0139119.
PR  16-JUN-1999;  99US-0139452.
PR  17-JUN-1999;  99US-0139492.
PR  18-JUN-1999;  99US-0139454.
PR  18-JUN-1999;  99US-0139455.
PR  18-JUN-1999;  99US-0139456.
PR  18-JUN-1999;  99US-0139457.
PR  18-JUN-1999;  99US-0139458.
PR  18-JUN-1999;  99US-0139459.
PR  18-JUN-1999;  99US-0139460.
PR  18-JUN-1999;  99US-0139461.
PR  18-JUN-1999;  99US-0139462.
PR  18-JUN-1999;  99US-0139463.
PR  17-JUN-1999;  99US-0139750.
PR  18-JUN-1999;  99US-0139763.
PR  21-JUN-1999;  99US-0139817.
PR  22-JUN-1999;  99US-0139899.
PR  23-JUN-1999;  99US-0140353.
PR  23-JUN-1999;  99US-0140354.
PR  24-JUN-1999;  99US-0140695.
PR  28-JUN-1999;  99US-0140823.
PR  29-JUN-1999;  99US-0140991.
PR  30-JUN-1999;  99US-0141287.
PR  01-JUL-1999;  99US-0141842.
PR  01-JUL-1999;  99US-0142154.
PR  02-JUL-1999;  99US-0142055.
PR  06-JUL-1999;  99US-0142390.
PR  08-JUL-1999;  99US-0142803.
PR  09-JUL-1999;  99US-0142920.
PR  12-JUL-1999;  99US-0142977.
PR  13-JUL-1999;  99US-0143542.
PR  14-JUL-1999;  99US-0143624.
PR  15-JUL-1999;  99US-0144005.
PR  16-JUL-1999;  99US-0144085.
PR  16-JUL-1999;  99US-0144086.
PR  19-JUL-1999;  99US-0144325.
PR  19-JUL-1999;  99US-0144331.
PR  19-JUL-1999;  99US-0144332.
PR  19-JUL-1999;  99US-0144333.

PR  19-JUL-1999;  99US-0144334.
PR  19-JUL-1999;  99US-0144335.
PR  20-JUL-1999;  99US-0144352.
PR  20-JUL-1999;  99US-0144632.
PR  20-JUL-1999;  99US-0144884.
PR  21-JUL-1999;  99US-0144814.
PR  21-JUL-1999;  99US-0145086.
PR  21-JUL-1999;  99US-0145088.
PR  22-JUL-1999;  99US-0145085.
PR  22-JUL-1999;  99US-0145087.
PR  22-JUL-1999;  99US-0145089.
PR  22-JUL-1999;  99US-0145192.
PR  23-JUL-1999;  99US-0145145.
PR  23-JUL-1999;  99US-0145218.
PR  26-JUL-1999;  99US-0145224.
PR  27-JUL-1999;  99US-0145276.
PR  27-JUL-1999;  99US-0145913.
PR  27-JUL-1999;  99US-0145918.
PR  28-JUL-1999;  99US-0145919.
PR  02-AUG-1999;  99US-0145951.
PR  02-AUG-1999;  99US-0146386.
PR  02-AUG-1999;  99US-0146388.
PR  03-AUG-1999;  99US-0146389.
PR  04-AUG-1999;  99US-0147038.
PR  04-AUG-1999;  99US-0147204.
PR  05-AUG-1999;  99US-0147302.
PR  05-AUG-1999;  99US-0147192.
PR  06-AUG-1999;  99US-0147260.
PR  06-AUG-1999;  99US-0147303.
PR  09-AUG-1999;  99US-0147416.
PR  09-AUG-1999;  99US-0147493.
PR  10-AUG-1999;  99US-0147935.
PR  11-AUG-1999;  99US-0148171.
PR  12-AUG-1999;  99US-0148341.
PR  13-AUG-1999;  99US-0148565.
PR  16-AUG-1999;  99US-0148684.
PR  17-AUG-1999;  99US-0149175.
PR  18-AUG-1999;  99US-0149368.
PR  18-AUG-1999;  99US-0149426.
PR  20-AUG-1999;  99US-0149722.
PR  20-AUG-1999;  99US-0149723.
PR  20-AUG-1999;  99US-0149929.
PR  23-AUG-1999;  99US-0149902.
PR  25-AUG-1999;  99US-0149930.
PR  25-AUG-1999;  99US-0150566.
PR  26-AUG-1999;  99US-0150884.
PR  27-AUG-1999;  99US-0151065.
PR  27-AUG-1999;  99US-0151066.
PR  30-AUG-1999;  99US-0151080.
PR  31-AUG-1999;  99US-0151303.
PR  01-SEP-1999;  99US-0151438.
PR  07-SEP-1999;  99US-0151930.
PR  10-SEP-1999;  99US-0152363.
PR  13-SEP-1999;  99US-0153070.
PR  15-SEP-1999;  99US-0153758.
PR  16-SEP-1999;  99US-0154018.
PR  20-SEP-1999;  99US-0154039.
PR  22-SEP-1999;  99US-0154779.
PR  23-SEP-1999;  99US-0155139.
PR  24-SEP-1999;  99US-0155486.
PR  28-SEP-1999;  99US-0155659.
PR  29-SEP-1999;  99US-0156458.
PR  04-OCT-1999;  99US-0156596.
PR  05-OCT-1999;  99US-0157117.
PR  06-OCT-1999;  99US-0157753.
PR  07-OCT-1999;  99US-0157865.
PR  08-OCT-1999;  99US-0158029.
PR  12-OCT-1999;  99US-0158232.
PR  13-OCT-1999;  99US-0158369.
PR  13-OCT-1999;  99US-0159293.
PR  13-OCT-1999;  99US-0159294.
PR  13-OCT-1999;  99US-0159295.
```

```
RESULT 18
AAG22264
ID   AAG22264 standard; Protein; 344 AA.
XX
AC   AAG22264;
XX
DT   17-OCT-2000  (first entry)
XX
DE   Arabidopsis thaliana protein fragment SEQ ID NO: 25125.
XX
KW   Protein identification; signal transduction pathway; metabolic pathway;
KW   hybridisation assay; genetic mapping; gene expression control; promoter;
KW   termination sequence.
XX
OS   Arabidopsis thaliana.
XX
PN   EP1033405-A2.
XX
PD   06-SEP-2000.
XX
PF   25-FEB-2000;  2000EP-0301439.
XX
PR   25-FEB-1999;  99US-0121825.
PR   05-MAR-1999;  99US-0123180.
PR   09-MAR-1999;  99US-0123548.
PR   23-MAR-1999;  99US-0125788.
PR   25-MAR-1999;  99US-0126264.
PR   29-MAR-1999;  99US-0126785.
PR   01-APR-1999;  99US-0127462.
PR   06-APR-1999;  99US-0128334.
PR   08-APR-1999;  99US-0128714.
PR   16-APR-1999;  99US-0129845.
PR   19-APR-1999;  99US-0130077.
PR   21-APR-1999;  99US-0130419.
PR   23-APR-1999;  99US-0130510.
PR   28-APR-1999;  99US-0130891.
PR   30-APR-1999;  99US-0131449.
PR   04-MAY-1999;  99US-0132407.
PR   05-MAY-1999;  99US-0132484.
PR   06-MAY-1999;  99US-0132485.
PR   06-MAY-1999;  99US-0132486.
PR   07-MAY-1999;  99US-0132487.
PR   07-MAY-1999;  99US-0132863.
PR   11-MAY-1999;  99US-0134256.
PR   14-MAY-1999;  99US-0134218.
PR   14-MAY-1999;  99US-0134219.
PR   14-MAY-1999;  99US-0134221.
PR   14-MAY-1999;  99US-0134370.
PR   18-MAY-1999;  99US-0134768.
PR   19-MAY-1999;  99US-0134941.
PR   20-MAY-1999;  99US-0133124.
PR   21-MAY-1999;  99US-0135353.
PR   24-MAY-1999;  99US-0135629.
PR   25-MAY-1999;  99US-0136021.
PR   27-MAY-1999;  99US-0136392.
PR   28-MAY-1999;  99US-0136782.
PR   01-JUN-1999;  99US-0137222.
PR   03-JUN-1999;  99US-0137528.
PR   04-JUN-1999;  99US-0137502.
PR   07-JUN-1999;  99US-0137724.
PR   08-JUN-1999;  99US-0138094.
PR   10-JUN-1999;  99US-0138540.
PR   10-JUN-1999;  99US-0138847.
PR   14-JUN-1999;  99US-0139119.
PR   16-JUN-1999;  99US-0139452.
PR   16-JUN-1999;  99US-0139453.
PR   16-JUN-1999;  99US-0139454.
PR   17-JUN-1999;  99US-0139455.
PR   18-JUN-1999;  99US-0139456.
PR   18-JUN-1999;  99US-0139457.
PR   18-JUN-1999;  99US-0139458.
PR   18-JUN-1999;  99US-0139459.
PR   18-JUN-1999;  99US-0139460.
PR   18-JUN-1999;  99US-0139461.
PR   18-JUN-1999;  99US-0139462.
PR   18-JUN-1999;  99US-0139463.
PR   18-JUN-1999;  99US-0139750.
PR   21-JUN-1999;  99US-0139817.
PR   22-JUN-1999;  99US-0139899.
PR   23-JUN-1999;  99US-0140353.
PR   23-JUN-1999;  99US-0140354.
PR   24-JUN-1999;  99US-0140695.
PR   28-JUN-1999;  99US-0140823.
PR   29-JUN-1999;  99US-0140991.
PR   30-JUN-1999;  99US-0141287.
PR   01-JUL-1999;  99US-0141842.
PR   01-JUL-1999;  99US-0142154.
PR   02-JUL-1999;  99US-0142055.
PR   14-OCT-1999;  99US-0159329.
PR   14-OCT-1999;  99US-0159330.
PR   14-OCT-1999;  99US-0159331.
PR   14-OCT-1999;  99US-0159637.
PR   14-OCT-1999;  99US-0159638.
PR   18-OCT-1999;  99US-0159584.
PR   21-OCT-1999;  99US-0160741.
PR   21-OCT-1999;  99US-0160767.
PR   21-OCT-1999;  99US-0160768.
PR   21-OCT-1999;  99US-0160770.
PR   21-OCT-1999;  99US-0160777.
PR   21-OCT-1999;  99US-0160814.
PR   21-OCT-1999;  99US-0160815.
PR   22-OCT-1999;  99US-0160980.
PR   22-OCT-1999;  99US-0160981.
PR   22-OCT-1999;  99US-0160989.
PR   25-OCT-1999;  99US-0161404.
PR   25-OCT-1999;  99US-0161405.
PR   25-OCT-1999;  99US-0161406.
PR   26-OCT-1999;  99US-0161359.
PR   26-OCT-1999;  99US-0161360.
PR   26-OCT-1999;  99US-0161361.
PR   28-OCT-1999;  99US-0161920.
PR   28-OCT-1999;  99US-0161992.
PR   28-OCT-1999;  99US-0161993.
PR   29-OCT-1999;  99US-0162142.
```

```
Query Match     33.8%;  Score 649.5;  DB 21;  Length 398;
Best Local Similarity   37.9%;  Pred. No. 5.3e-61;
Matches 134;  Conservative 71;  Mismatches 88;  Indels 61;  Gaps 4;

Qy    60 DSHCDPKKGKAQRQLYVASAICLLFMIGEVVGGGYLAHSLAVMTDAAHLLTDFASMLISLF 119
             |:|      |:|  :|:||   ||||  :: |||||:|: ||||:|| |:  |: |||| ||:
Db    47 DAH---ERSASMRKLCIAVVLCLVFMSVEVVGGIKANSLAILTDAAHLLSDVAAFAISLF 103

Qy   120 SLWMSSRPATKTMNFGWQRAELLGALVSVLSIWVVTGVLVLYLAVERLISGDYEIDGGTML 179
             ||        |  ||||:||||||||||||  | | |||:|| |:||||| |  |:|   |
Db   104 SLWAAGWEATPRQTYGFFRIEILGALVSIQLIWLLTGILVYEAIIRIVTETSEVNGFLMF 163

Qy   180 ITSGCAVAVNIIMGLTL-----HQSGHGHSHG----------------------- 209
             :  ||||||| :|  |      |:| |||||
Db   164 LVAAFGLVVNIIMAVLLGHDHGHSHGHGHDHHHNHSHGYVTTHHHHHDHEHGHSHG 223

Qy   210 --------------QOENPSVRAAFIHVIGDFMQSMGVLVAAYIL 241
                           |:: |:| |||  ||:|:||  |  |:
Db   224 HGEDKHHAHGDVTEQLLDKSKTQVAAAKEKRKRNINLQGAYLHVLGDSIQSVGVMIGGAII 283

Qy   242 YFKPEYKYVDPICTFVFSILVLGTTLILLRDVILVLMEGTPKGVDFTAVRDLLLSVEGVE 301
             ::|||:|| |||||:|||:|||  :|||||  |: ||| || |||||  | |:  || |
Db   284 WYNPEWKIVDLICTLAFSVIVLGTTINMIRNILEVLMESTPREIDATKLEKGLLEMEEVV 343

Qy   302 ALHSLHIWALTVAQPVLSVHIAIAONTDAQAVLKTASSRLQGKFHFHTVTIQIE 355
             |:|| ||||:|:|||: :|: : |  :|  |  :  |:||:|| ||:| |||:|
Db   344 AVHELHIWAITVGKVLLACHVNIRPEADADMVLNKVIDYIRREYNISHVTIQIE 397
```

PR   06-JUL-1999;   99US-0142390.
PR   08-JUL-1999;   99US-0142803.
PR   09-JUL-1999;   99US-0142920.
PR   12-JUL-1999;   99US-0142977.
PR   13-JUL-1999;   99US-0143542.
PR   14-JUL-1999;   99US-0143624.
PR   15-JUL-1999;   99US-0144005.
PR   16-JUL-1999;   99US-0144085.
PR   16-JUL-1999;   99US-0144086.
PR   19-JUL-1999;   99US-0144252.
PR   19-JUL-1999;   99US-0144325.
PR   19-JUL-1999;   99US-0144331.
PR   19-JUL-1999;   99US-0144332.
PR   19-JUL-1999;   99US-0144333.
PR   19-JUL-1999;   99US-0144334.
PR   19-JUL-1999;   99US-0144335.
PR   20-JUL-1999;   99US-0144352.
PR   20-JUL-1999;   99US-0144632.
PR   20-JUL-1999;   99US-0145087.
PR   20-JUL-1999;   99US-0145089.
PR   22-JUL-1999;   99US-0145192.
PR   22-JUL-1999;   99US-0145276.
PR   23-JUL-1999;   99US-0145145.
PR   23-JUL-1999;   99US-0144814.
PR   23-JUL-1999;   99US-0144884.
PR   26-JUL-1999;   99US-0145913.
PR   27-JUL-1999;   99US-0145518.
PR   27-JUL-1999;   99US-0145224.
PR   28-JUL-1999;   99US-0145951.
PR   02-AUG-1999;   99US-0146386.
PR   02-AUG-1999;   99US-0146388.
PR   02-AUG-1999;   99US-0146389.
PR   03-AUG-1999;   99US-0147038.
PR   04-AUG-1999;   99US-0147204.
PR   04-AUG-1999;   99US-0147302.
PR   05-AUG-1999;   99US-0147192.
PR   05-AUG-1999;   99US-0147260.
PR   06-AUG-1999;   99US-0147303.
PR   06-AUG-1999;   99US-0147416.
PR   09-AUG-1999;   99US-0147493.
PR   09-AUG-1999;   99US-0147935.
PR   10-AUG-1999;   99US-0148171.
PR   11-AUG-1999;   99US-0148319.
PR   12-AUG-1999;   99US-0148341.
PR   13-AUG-1999;   99US-0148565.
PR   13-AUG-1999;   99US-0148684.
PR   16-AUG-1999;   99US-0149368.
PR   17-AUG-1999;   99US-0149175.
PR   18-AUG-1999;   99US-0149426.
PR   20-AUG-1999;   99US-0149722.
PR   20-AUG-1999;   99US-0149723.
PR   20-AUG-1999;   99US-0149929.
PR   23-AUG-1999;   99US-0149902.
PR   23-AUG-1999;   99US-0149930.
PR   25-AUG-1999;   99US-0150556.
PR   26-AUG-1999;   99US-0150884.
PR   27-AUG-1999;   99US-0151065.
PR   27-AUG-1999;   99US-0151066.
PR   27-AUG-1999;   99US-0151080.
PR   30-AUG-1999;   99US-0151303.
PR   31-AUG-1999;   99US-0151438.
PR   01-SEP-1999;   99US-0151930.
PR   01-SEP-1999;   99US-0152363.
PR   07-SEP-1999;   99US-0153070.
PR   10-SEP-1999;   99US-0153070.
PR   13-SEP-1999;   99US-0153758.
PR   15-SEP-1999;   99US-0154018.
PR   16-SEP-1999;   99US-0154039.
PR   20-SEP-1999;   99US-0154779.
PR   20-SEP-1999;   99US-0154779.
PR   22-SEP-1999;   99US-0155139.

PR   23-SEP-1999;   99US-0155486.
PR   24-SEP-1999;   99US-0155659.
PR   28-SEP-1999;   99US-0156458.
PR   29-SEP-1999;   99US-0156596.
PR   04-OCT-1999;   99US-0157117.
PR   05-OCT-1999;   99US-0157753.
PR   06-OCT-1999;   99US-0157865.
PR   07-OCT-1999;   99US-0158029.
PR   08-OCT-1999;   99US-0158232.
PR   12-OCT-1999;   99US-0158369.
PR   13-OCT-1999;   99US-0159293.
PR   13-OCT-1999;   99US-0159294.
PR   13-OCT-1999;   99US-0159295.
PR   14-OCT-1999;   99US-0159329.
PR   14-OCT-1999;   99US-0159330.
PR   14-OCT-1999;   99US-0159331.
PR   14-OCT-1999;   99US-0159637.
PR   14-OCT-1999;   99US-0159638.
PR   18-OCT-1999;   99US-0159584.
PR   21-OCT-1999;   99US-0160815.
PR   21-OCT-1999;   99US-0160814.
PR   21-OCT-1999;   99US-0160770.
PR   21-OCT-1999;   99US-0160768.
PR   21-OCT-1999;   99US-0160767.
PR   22-OCT-1999;   99US-0160980.
PR   22-OCT-1999;   99US-0160981.
PR   22-OCT-1999;   99US-0160989.
PR   25-OCT-1999;   99US-0161404.
PR   25-OCT-1999;   99US-0161405.
PR   25-OCT-1999;   99US-0161406.
PR   26-OCT-1999;   99US-0161359.
PR   26-OCT-1999;   99US-0161360.
PR   26-OCT-1999;   99US-0161361.
PR   28-OCT-1999;   99US-0161920.
PR   28-OCT-1999;   99US-0161992.
PR   28-OCT-1999;   99US-0161993.
PR   29-OCT-1999;   99US-0162142.

```
Query Match     33.4%;     Score 643; DB 21;   Length 344;
Best Local Similarity   38.6%;   Pred. No. 2.1e-60;
Matches  132;  Conservative  67;  Mismatches  85;  Indels  58;  Gaps  3;

Qy  72 RQLYVASAICLLPMIGEVGVGYLAHSLAVMTDAAHLLTDFASMLISLFSLWMSSRPAIKT 131
        |:| | |  ||:|| |||||| ::||||||||||||:: || |||||||||||      |
Db   2 RKLCIAVVLCLVFMSVEVGGIKANSLAILTDAAHLLSDVAAFAISLFSLWAAGWEATPR  61

Qy 132 MNFGWQPRAEILGALVSVLSIWVTGVLVYLAYERLISGDYEIDGGTMLITSCCAVAVNII 191
        ::| :| | ||||||||: |||:|||||| :  ||  | :|   |:|| || |||||||
Db  62 QYTGFFRIEILGALVSIQLIWILITGILVYEAIIRIVTETSEVNGFLMPLVAAPGLVVNII 121

Qy 192 MGLTL-----HQSGHGHSHG------------TTN------ 209
        ||:|        |||||||              ||:
Db 122 MAVLLGHDHGHSHGHGHGHDHHNHSHGGVTVTTHHHHHHDEHGHSHGHGHGEDKHHAHGDV 181

Qy 210 ------------QQEENPSVRAAFIHVIGDFMQSMGVLVAAYILYFKPEYKIYVDPI 253
                    ::: |  :|:|  |||:||:|| ::|::| ||||||:  |  || |
Db 182 TEQLLDKSKTQVAAKEKKRNINLQGAYLHVLGDSIQSVGVMIGGAIIWYNPEWKIVDLI 241

Qy 254 CTFVFSILIVLLGTTLTLTLRDVILVLMEGTPKGVDFTAVRDLLLSVEGVEALHSLHIWALTV 313
        || ||| ||| |||| | :  | || || |||. |   |      ||  |  |:|:|||||:||
Db 242 CTLAFSVIVLGTTINMIRNILEVLMESTPREIDATKLEKGLLEMEEVVAVHELHIWAITV 301

Qy 314 AQPVLSVHIAIAQNTDAQAVLKTASSRLQGKFHFHTVTIQIE 355
        ::: :| :  :|       :| |:    ::|   ||||||||
Db 302 GKVLLACHVNIRPEADADMVLNKVIDYIRREYNISHVYTIQIE 343
```

RESULT 19
AAG43479
ID   AAG43479 standard; Protein; 344 AA.
XX
AC   AAG43479;

```
XX
DT   18-OCT-2000  (first entry)
XX
DE   Arabidopsis thaliana protein fragment SEQ ID NO: 54350.
XX
KW   Protein identification; signal transduction pathway; metabolic pathway;
KW   hybridisation assay; genetic mapping; gene expression control; promoter;
KW   termination sequence.
XX
OS   Arabidopsis thaliana.
XX
PN   EP1033405-A2.
XX
PD   06-SEP-2000.
XX
PF   25-FEB-2000; 2000EP-0301439.

PR   25-FEB-1999; 99US-0121825.        PR   18-JUN-1999; 99US-0137750.
PR   05-MAR-1999; 99US-0123180.        PR   18-JUN-1999; 99US-0139763.
PR   09-MAR-1999; 99US-0123548.        PR   21-JUN-1999; 99US-0139817.
PR   23-MAR-1999; 99US-0125788.        PR   22-JUN-1999; 99US-0139899.
PR   25-MAR-1999; 99US-0126264.        PR   23-JUN-1999; 99US-0140353.
PR   29-MAR-1999; 99US-0126785.        PR   23-JUN-1999; 99US-0140354.
PR   01-APR-1999; 99US-0127462.        PR   24-JUN-1999; 99US-0140695.
PR   06-APR-1999; 99US-0128234.        PR   28-JUN-1999; 99US-0140823.
PR   08-APR-1999; 99US-0128714.        PR   29-JUN-1999; 99US-0140991.
PR   16-APR-1999; 99US-0129845.        PR   30-JUN-1999; 99US-0141287.
PR   19-APR-1999; 99US-0130077.        PR   01-JUL-1999; 99US-0141842.
PR   21-APR-1999; 99US-0130449.        PR   01-JUL-1999; 99US-0142154.
PR   23-APR-1999; 99US-0130510.        PR   02-JUL-1999; 99US-0142055.
PR   23-APR-1999; 99US-0130891.        PR   06-JUL-1999; 99US-0142290.
PR   28-APR-1999; 99US-0131449.        PR   08-JUL-1999; 99US-0142803.
PR   30-APR-1999; 99US-0132048.        PR   09-JUL-1999; 99US-0142920.
PR   30-APR-1999; 99US-0132407.        PR   12-JUL-1999; 99US-0142977.
PR   04-MAY-1999; 99US-0132484.        PR   13-JUL-1999; 99US-0143542.
PR   05-MAY-1999; 99US-0132485.        PR   14-JUL-1999; 99US-0143624.
PR   06-MAY-1999; 99US-0132486.        PR   15-JUL-1999; 99US-0144005.
PR   06-MAY-1999; 99US-0132487.        PR   16-JUL-1999; 99US-0144085.
PR   07-MAY-1999; 99US-0132863.        PR   16-JUL-1999; 99US-0144086.
PR   11-MAY-1999; 99US-0134256.        PR   19-JUL-1999; 99US-0144325.
PR   14-MAY-1999; 99US-0134218.        PR   19-JUL-1999; 99US-0144331.
PR   14-MAY-1999; 99US-0134219.        PR   19-JUL-1999; 99US-0144332.
PR   14-MAY-1999; 99US-0134221.        PR   19-JUL-1999; 99US-0144333.
PR   18-MAY-1999; 99US-0134370.        PR   19-JUL-1999; 99US-0144334.
PR   18-MAY-1999; 99US-0134768.        PR   19-JUL-1999; 99US-0144335.
PR   19-MAY-1999; 99US-0134941.        PR   20-JUL-1999; 99US-0144352.
PR   20-MAY-1999; 99US-0135124.        PR   20-JUL-1999; 99US-0144632.
PR   21-MAY-1999; 99US-0135353.        PR   20-JUL-1999; 99US-0144884.
PR   24-MAY-1999; 99US-0135629.        PR   21-JUL-1999; 99US-0144814.
PR   25-MAY-1999; 99US-0136021.        PR   21-JUL-1999; 99US-0145086.
PR   27-MAY-1999; 99US-0136392.        PR   21-JUL-1999; 99US-0145086.
PR   28-MAY-1999; 99US-0136782.        PR   22-JUL-1999; 99US-0145087.
PR   01-JUN-1999; 99US-0137222.        PR   22-JUL-1999; 99US-0145085.
PR   03-JUN-1999; 99US-0137528.        PR   22-JUL-1999; 99US-0145192.
PR   04-JUN-1999; 99US-0137502.        PR   23-JUL-1999; 99US-0145145.
PR   07-JUN-1999; 99US-0137724.        PR   23-JUL-1999; 99US-0145218.
PR   08-JUN-1999; 99US-0138094.        PR   23-JUL-1999; 99US-0145224.
PR   10-JUN-1999; 99US-0138540.        PR   26-JUL-1999; 99US-0145276.
PR   14-JUN-1999; 99US-0138847.        PR   27-JUL-1999; 99US-0145913.
PR   16-JUN-1999; 99US-0139119.        PR   27-JUL-1999; 99US-0145518.
PR   16-JUN-1999; 99US-0139452.        PR   27-JUL-1999; 99US-0145919.
PR   17-JUN-1999; 99US-0139453.        PR   28-JUL-1999; 99US-0145951.
PR   18-JUN-1999; 99US-0139454.        PR   02-AUG-1999; 99US-0145386.
PR   18-JUN-1999; 99US-0139455.        PR   02-AUG-1999; 99US-0146388.
PR   18-JUN-1999; 99US-0139456.        PR   02-AUG-1999; 99US-0146389.
PR   18-JUN-1999; 99US-0139457.        PR   03-AUG-1999; 99US-0147038.
PR   18-JUN-1999; 99US-0139458.        PR   04-AUG-1999; 99US-0147204.
PR   18-JUN-1999; 99US-0139459.        PR   05-AUG-1999; 99US-0147192.
PR   18-JUN-1999; 99US-0139460.        PR   05-AUG-1999; 99US-0147260.
PR   18-JUN-1999; 99US-0139461.        PR   06-AUG-1999; 99US-0147303.
PR   18-JUN-1999; 99US-0139462.        PR   06-AUG-1999; 99US-0147416.
PR   18-JUN-1999; 99US-0139463.        PR   09-AUG-1999; 99US-0147493.
                                       PR   09-AUG-1999; 99US-0147935.
                                       PR   10-AUG-1999; 99US-0148171.
                                       PR   11-AUG-1999; 99US-0148819.
                                       PR   12-AUG-1999; 99US-0148341.
                                       PR   13-AUG-1999; 99US-0148565.
                                       PR   13-AUG-1999; 99US-0148684.
                                       PR   16-AUG-1999; 99US-0149368.
                                       PR   17-AUG-1999; 99US-0149175.
                                       PR   18-AUG-1999; 99US-0149426.
                                       PR   20-AUG-1999; 99US-0149722.
                                       PR   20-AUG-1999; 99US-0149723.
                                       PR   20-AUG-1999; 99US-0149929.
                                       PR   23-AUG-1999; 99US-0149902.
                                       PR   23-AUG-1999; 99US-0149930.
                                       PR   25-AUG-1999; 99US-0150566.
                                       PR   26-AUG-1999; 99US-0150884.
```

```
PR  27-AUG-1999;  99US-0151065.
PR  27-AUG-1999;  99US-0151066.
PR  27-AUG-1999;  99US-0151080.
PR  30-AUG-1999;  99US-0151303.
PR  31-AUG-1999;  99US-0151438.
PR  01-SEP-1999;  99US-0151930.
PR  07-SEP-1999;  99US-0152363.
PR  10-SEP-1999;  99US-0153070.
PR  13-SEP-1999;  99US-0153758.
PR  15-SEP-1999;  99US-0154018.
PR  16-SEP-1999;  99US-0154039.
PR  20-SEP-1999;  99US-0154779.
PR  22-SEP-1999;  99US-0155139.
PR  23-SEP-1999;  99US-0155486.
PR  24-SEP-1999;  99US-0155659.
PR  28-SEP-1999;  99US-0156458.
PR  29-SEP-1999;  99US-0156996.
PR  04-OCT-1999;  99US-0157117.
PR  05-OCT-1999;  99US-0157753.
PR  06-OCT-1999;  99US-0157865.
PR  07-OCT-1999;  99US-0158029.
PR  08-OCT-1999;  99US-0158232.
PR  12-OCT-1999;  99US-0158369.
PR  13-OCT-1999;  99US-0159293.
PR  13-OCT-1999;  99US-0159294.
PR  13-OCT-1999;  99US-0159295.
PR  14-OCT-1999;  99US-0159329.
PR  14-OCT-1999;  99US-0159330.
PR  14-OCT-1999;  99US-0159331.
PR  14-OCT-1999;  99US-0159637.
PR  14-OCT-1999;  99US-0159638.
PR  18-OCT-1999;  99US-0159584.
PR  21-OCT-1999;  99US-0160741.
PR  21-OCT-1999;  99US-0160768.
PR  21-OCT-1999;  99US-0160770.
PR  21-OCT-1999;  99US-0160815.
PR  22-OCT-1999;  99US-0160767.
PR  22-OCT-1999;  99US-0160814.
PR  22-OCT-1999;  99US-0160980.
PR  22-OCT-1999;  99US-0160981.
PR  22-OCT-1999;  99US-0160989.
PR  25-OCT-1999;  99US-0161404.
PR  25-OCT-1999;  99US-0161405.
PR  25-OCT-1999;  99US-0161406.
PR  26-OCT-1999;  99US-0161359.
PR  26-OCT-1999;  99US-0161360.
PR  26-OCT-1999;  99US-0161361.
PR  28-OCT-1999;  99US-0161902.
PR  28-OCT-1999;  99US-0161920.
PR  28-OCT-1999;  99US-0161992.
PR  28-OCT-1999;  99US-0161993.
PR  29-OCT-1999;  99US-0162142.
```

Query Match   Score 643;   DB 21; Length 344;
Best Local Similarity   33.4%;   Pred. No. 2.1e-60;
Matches 132; Conservative   38.6%;   Mismatches 85;   Indels 58;   Gaps 3;
                                             67;

```
Qy   72  RQLYVASACILLFMIGEVVGGYLAHSLAVWTDAAHLLTDFASMLISLFSLWMSSRPATKT  131
             |:|  ||::||||| ::|.|::: |||| ||:|||||||||||| :|  |||||||||
Db    2  RKLCIAVVLCLVFMSVEVVGGIKANSLAILTDAAHLLSDVAAFAISLFSLWAAGWEATPR   61
```

```
Qy  132  MNFGWQORAEILGALVSVLSIWVVTGVLVYLAVERLISGDYEIDGGTMLITSGCAVAVNII  191
           :|:   |||||||| :|||||||||| |: :|||| ||||||||||||||||||||||||
Db   62  QTYGFFRIELGALVSIQLIWLLTGILVYEAIIRIVTETSEVNGELMFLVAAFGLVVNII    121
```

```
Qy  192  MGLTL------HOSGHHSHG-----------                               209
           ||:| |      | |||| ||
Db  122  MAVLLGHDHGHSHGHGHGHDHHHHSHGVTVTTHHHHDHEHGHSHGHGHGEDKHHAHGDV   181
```

```
Qy  210  ------------------QEEENPSVRAAFIHVIGDFMQSMGVLVAAVILYFKPEYKYVDPI  253
                              ||||| ||| |||| ||
Db  182  TEQLLDKSKTQVAAAKEKRKRNINLQGAYLHVLGDSIQSVGVMIGGAIIWYNPEWKIVDLI  241
```

```
Qy  254  CTFVPSILVLGTTLTLTLRDVILVLMEGTPKGVDFTAVRDLLLSVEGVEALHSLHIWALTV  313
```

```
Db  242  CTIAFSVTVLGTTPINMIRNILEVLMESTPREIDATKLEKGLLEMEEVVAVHELHIWALTV  301
             ||| ||::|||||    :| ::|   ||| :|||:|   |||||||||  ||| :||||| :||
Qy  314  AQPVLSVHIAIAQNTDAQAVLKTASSRLQGKFHFHTVTIQIE                    355
             :|||  |:|    ||  :||  | |:| |   |  |||||
Db  302  GKVLLACHVNIRPEADADMVLNKVIDYIRREYNISHVTIQIE                    343
```

RESULT 20
AAE09323
ID   AAE09323 standard;   Protein;   349 AA.
XX
AC   AAE09323;
XX
DT   22-NOV-2001   (first entry)
XX
DE   Soybean zinc transporter (ZnT-2) protein from clone se6.pk0012.h2:fis.
XX
KW   Heavy metal transporter; zinc transporter; transgenic plant;
KW   soybean; ZnT-2.
XX
OS   Glycine max.
XX
OS   Glycine max.
XX
RN   US6278042-B1.
PN   US6278042-B1.
XX
PD   21-AUG-2001.
XX
PF   14-DEC-1999;   99US-0461474.
XX
PR   16-DEC-1998;   98US-0112562.
XX
PA   (DUPO) DU PONT DE NEMOURS & CO E I.
XX
PI   Allen SM, Rafalski JA, Sakai H;
XX
DR   WPI; 2001-540412/60.
DR   N-PSDB; AAD16067.
XX
PT   New arsenic transporter polypeptides and polynucleotides, useful for
PT   producing transgenic plants with altered level of heavy metal tolerance
-
XX
CC   Example 4; Fig 2; 31pp; English.
CC
CC   The invention relates to heavy metal (e.g. arsenite and zinc)
CC   transporter polypeptides and polynucleotides. The polypeptides of
CC   the invention are useful for producing antibodies that are useful
CC   for screening cDNA expression libraries to isolate full length
CC   clones of interest. The antibodies are also useful for detecting
CC   the polypeptides in situ, in cells or in vitro, in cell extracts.
CC   Nucleic acid molecules of the invention are useful for producing
CC   transgenic plants with altered tolerance to heavy metals.They are
CC   also useful as probes for genetically and physically mapping the
CC   genes that are a part of, and as markers for traits linked to those
CC   genes. The present sequence is soybean zinc transporter (ZnT-2)
CC   protein.
XX
SQ   Sequence   349 AA;

Query Match   Score 613.5;   DB 22; Length 349;
Best Local Similarity   31.9%;   Pred. No. 3.3e-57;
Matches 128; Conservative   37.2%;   Mismatches 81;   Indels 73;   Gaps 3;
                                             62;

```
Qy   85  MIGEVVGGYLAHSLAVWTDAAHLLTDFASMLISLFSLWMSSRPATKTMNFGWQORAEILGA  144
            |||||| :|| |||:| ||||||| ::|| |||||||:  ||  ||:|   ||||:||  |
Db    5  MTVEVVGGIKANSLAILTDAAHLLSDVASFAISLFSLWAAGWEATPRQSYGFFRIELGA    64
```

```
Qy  145  LVSVLSIWVVTGVLVYLAVERLISGDYEIDGGTMLITSGCAVAVNIIMGLTL------HQ-  198
           ||::||||||| || | |:|||::|:|||| |||||||||||||||||||| |:     | |
Db   65  LVSIQMIWLLAGILVYEAIIDRIIAGPKNVDGFLMFLVSAFGLVVNIIMALLGHDHGHRH  124
```

```
Qy  199  ------------SGHGHSH------------------------                    205
                       ||||||
```

```
Db   125  AGHSHGHGHDGHGHSHGFTMSTHCDAKHTKDQHHHHHHHDENHPKDAHHHTDEDHLHHHA  184
          |||   ||     ||  ||    ||||  | :| | :||
Qy   206  ------------------GTTNQEENPSVRAAFIHVIGDFMQSMGVLVAAYILYFKPEYKYVD  251
          ||  :| | :||  || :::|  |  :|| :: || |
Db   185  HKEVTELLLGESKGGGTRKKKQWNINVQGAYLHVLGDSIQSIGVMIGGAVIWNPRWQIVD  244
          |||:||:|||| ::::|||  ::|| | | :|||||
Qy   252  PICTFVFSILVLGTTLFLLRDVILVLMEGTPKGVDFTAVRDLLLSVEGVEALHSLHIWAL  311
          |:| ||:||:|| :||||| ::||||| ||| || ||| :|||:| |||||||
Db   245  LICTLIFSVIVWGTTINMLRNILEVLMENTPREIDATKLERGLLDMEDVVAVHELHIWAI  304
          ||  ||  | :  :: ||||||
Qy   312  TVAQPVLSVHIAIAQNTDAQAVLKTASSRLQGKFHFHTVTIQIE  355
          ||  || ||:::|   ::   ||||||
Db   305  TVGKVLLACHVKIRREADADLVLDKVIDYIKRVYNISHVTIQIE  348
```

RESULT 21

AAG22265

ID   AAG22265 standard; Protein; 330 AA.
XX
AC   AAG22265;
XX
DT   17-OCT-2000 (first entry)
XX
DE   Arabidopsis thaliana protein fragment SEQ ID NO: 25126.
XX
KW   Protein identification; signal transduction pathway; metabolic pathway;
KW   hybridisation assay; genetic mapping; gene expression control; promoter;
KW   termination sequence.
XX
OS   Arabidopsis thaliana.
XX
PN   EP1033405-A2.
XX
PD   06-SEP-2000.
XX
PF   25-FEB-2000; 2000EP-0301439.
XX

PR   25-FEB-1999; 99US-0121825.
PR   05-MAR-1999; 99US-0123180.
PR   09-MAR-1999; 99US-0123548.
PR   23-MAR-1999; 99US-0125788.
PR   25-MAR-1999; 99US-0126264.
PR   29-MAR-1999; 99US-0126785.
PR   01-APR-1999; 99US-0127462.
PR   06-APR-1999; 99US-0128234.
PR   08-APR-1999; 99US-0128714.
PR   16-APR-1999; 99US-0129845.
PR   19-APR-1999; 99US-0130077.
PR   21-APR-1999; 99US-0130449.
PR   23-APR-1999; 99US-0130510.
PR   23-APR-1999; 99US-0130891.
PR   28-APR-1999; 99US-0131449.
PR   30-APR-1999; 99US-0132048.
PR   30-APR-1999; 99US-0132407.
PR   04-MAY-1999; 99US-0132484.
PR   05-MAY-1999; 99US-0132485.
PR   06-MAY-1999; 99US-0132486.
PR   06-MAY-1999; 99US-0132487.
PR   07-MAY-1999; 99US-0132863.
PR   11-MAY-1999; 99US-0134256.
PR   14-MAY-1999; 99US-0134218.
PR   14-MAY-1999; 99US-0134219.
PR   14-MAY-1999; 99US-0134221.
PR   18-MAY-1999; 99US-0134370.
PR   18-MAY-1999; 99US-0134768.
PR   19-MAY-1999; 99US-0134941.
PR   20-MAY-1999; 99US-0135124.
PR   21-MAY-1999; 99US-0135353.
PR   24-MAY-1999; 99US-0135629.
PR   25-MAY-1999; 99US-0136021.
PR   27-MAY-1999; 99US-0136392.
PR   28-MAY-1999; 99US-0136782.
PR   01-JUN-1999; 99US-0137222.

PR   03-JUN-1999; 99US-0137528.
PR   04-JUN-1999; 99US-0137502.
PR   07-JUN-1999; 99US-0137724.
PR   08-JUN-1999; 99US-0138094.
PR   10-JUN-1999; 99US-0138540.
PR   10-JUN-1999; 99US-0138847.
PR   14-JUN-1999; 99US-0139119.
PR   16-JUN-1999; 99US-0139452.
PR   16-JUN-1999; 99US-0139453.
PR   17-JUN-1999; 99US-0139454.
PR   18-JUN-1999; 99US-0139456.
PR   18-JUN-1999; 99US-0139457.
PR   18-JUN-1999; 99US-0139458.
PR   18-JUN-1999; 99US-0139459.
PR   18-JUN-1999; 99US-0139460.
PR   18-JUN-1999; 99US-0139461.
PR   18-JUN-1999; 99US-0139462.
PR   18-JUN-1999; 99US-0139463.
PR   18-JUN-1999; 99US-0139750.
PR   18-JUN-1999; 99US-0139763.
PR   21-JUN-1999; 99US-0139817.
PR   22-JUN-1999; 99US-0139899.
PR   23-JUN-1999; 99US-0140353.
PR   23-JUN-1999; 99US-0140354.
PR   24-JUN-1999; 99US-0140695.
PR   28-JUN-1999; 99US-0140823.
PR   29-JUN-1999; 99US-0140991.
PR   30-JUN-1999; 99US-0141287.
PR   01-JUL-1999; 99US-0141842.
PR   01-JUL-1999; 99US-0142154.
PR   02-JUL-1999; 99US-0142055.
PR   06-JUL-1999; 99US-0142390.
PR   08-JUL-1999; 99US-0142803.
PR   09-JUL-1999; 99US-0142920.
PR   12-JUL-1999; 99US-0142977.
PR   13-JUL-1999; 99US-0143542.
PR   14-JUL-1999; 99US-0143624.
PR   15-JUL-1999; 99US-0144005.
PR   16-JUL-1999; 99US-0144085.
PR   16-JUL-1999; 99US-0144086.
PR   19-JUL-1999; 99US-0144325.
PR   19-JUL-1999; 99US-0144331.
PR   19-JUL-1999; 99US-0144332.
PR   19-JUL-1999; 99US-0144333.
PR   19-JUL-1999; 99US-0144334.
PR   19-JUL-1999; 99US-0144335.
PR   20-JUL-1999; 99US-0144352.
PR   20-JUL-1999; 99US-0144632.
PR   20-JUL-1999; 99US-0144804.
PR   21-JUL-1999; 99US-0144814.
PR   21-JUL-1999; 99US-0145086.
PR   21-JUL-1999; 99US-0145087.
PR   22-JUL-1999; 99US-0145085.
PR   22-JUL-1999; 99US-0145089.
PR   22-JUL-1999; 99US-0145145.
PR   23-JUL-1999; 99US-0145218.
PR   23-JUL-1999; 99US-0145224.
PR   26-JUL-1999; 99US-0145276.
PR   27-JUL-1999; 99US-0145913.
PR   27-JUL-1999; 99US-0145918.
PR   28-JUL-1999; 99US-0145519.
PR   02-AUG-1999; 99US-0145951.
PR   02-AUG-1999; 99US-0146386.
PR   02-AUG-1999; 99US-0146388.
PR   03-AUG-1999; 99US-0146389.
PR   03-AUG-1999; 99US-0147038.
PR   04-AUG-1999; 99US-0147204.
PR   04-AUG-1999; 99US-0147302.
PR   05-AUG-1999; 99US-0147192.
```

```
Query Match   31.6%;   Score 608;   DB 21;   Length 330;

Best Local Similarity   38.3%;   Conservative 63;   Pred. No. 1.2e-56;
Matches 125;   Mismatches 80;   Indels 58;   Gaps 3;

Qy   88  EVVGGYLAHSLAVMTDAAHLLTDFASMLISLFSLWMSSRPATKTMNFGWQRAEILGALVS  147
         |||| | :|:||::.:|||||||:|   |  ||||||||| : :|.|:|||||||||
Db    4  EVVGGIKANSLAILTDAAHLLSDVAAFAISLFSLWAAGWEATPRQTYGFFRIEILGALVS   63

Qy  148  VLSIWVVTGVLVYLAVERLISGDYEIDGGTMLITSGCAVAVNIIMGLTL------HQSGHG  202
         :|::|::|| ||| :: :|: :  :|:: |::|::| |:|||||| |      |||
Db   64  IQLIWLLTGILVYEAIIRIVTETSEVNGFLMFLVAAFGLVVNIIMAVLLGHDHGHSHGHG  123

Qy  203  HSHG--------------TN-----------------------------------  209
         ||:|              |:
Db  124  HGHGDHDHNHSHGVTVTHTHHHHDHEHGHSHGHGEDKHHAHGDVTEQLLDKSKTQVAAKE  183

Qy  210  QQEENPSVRAAFIHVIGDFMQSMGVLVAAYILYFKPEYKYVDPICTFVFSILVLGTTLTI  269
         ::|:. :|:::|:|::||.:||:||::| |:  ||: ::::|::|||.||:::|||||::
Db  184  KRKRNINLQGAYLHVLGDSIQSVGVMIGGAIIWYNPEWKIVDLICTLAFSVIVLGTTINM  243

Qy  270  LRDVILVLMEGTPKGVDFTAVRDLLLSVEGVEALHSLIHWALTVAQPVLSVHIAIAQNTD  329
         :|::| |||.:|::::|:| :  :| : :.::|||:||.|:|||| ::|:.|:  :  :|
Db  244  IRNILEVLMESTPREIDATKLEKGLLEMEEVVAVHELHIWAITVGKVLLACHVNIRPEAD  303

Qy  330  AQAVLKTASSRLQGKFHFHTVTIQIE  355
         | :||: .  :: : :::||||||||
Db  304  ADMVLNKVIDYIRREYNISHVTIQIE  329


RESULT 22
AAG43480
ID   AAG43480 standard; protein;  330 AA.
XX
AC   AAG43480;
XX
DT   18-OCT-2000  (first entry)
XX
DE   Arabidopsis thaliana protein fragment SEQ ID NO: 54351.
XX
KW   Protein identification; signal transduction pathway; metabolic pathway;
KW   hybridisation assay; genetic mapping; gene expression control; promoter;
KW   termination sequence.
XX
OS   Arabidopsis thaliana.
XX
PN   EP1033405-A2.
XX
PD   06-SEP-2000.
XX
PF   25-FEB-2000; 2000EP-0301439.
XX
```

PR   05-AUG-1999; 99US-0147260.
PR   06-AUG-1999; 99US-0147303.
PR   06-AUG-1999; 99US-0147416.
PR   09-AUG-1999; 99US-0147493.
PR   09-AUG-1999; 99US-0147935.
PR   10-AUG-1999; 99US-0148171.
PR   11-AUG-1999; 99US-0148319.
PR   12-AUG-1999; 99US-0148341.
PR   13-AUG-1999; 99US-0148565.
PR   13-AUG-1999; 99US-0148684.
PR   16-AUG-1999; 99US-0149368.
PR   17-AUG-1999; 99US-0149175.
PR   18-AUG-1999; 99US-0149426.
PR   20-AUG-1999; 99US-0149722.
PR   20-AUG-1999; 99US-0149723.
PR   20-AUG-1999; 99US-0149929.
PR   23-AUG-1999; 99US-0149902.
PR   23-AUG-1999; 99US-0149930.
PR   25-AUG-1999; 99US-0150566.
PR   26-AUG-1999; 99US-0150884.
PR   27-AUG-1999; 99US-0151065.
PR   27-AUG-1999; 99US-0151066.
PR   27-AUG-1999; 99US-0151080.
PR   30-AUG-1999; 99US-0151103.
PR   31-AUG-1999; 99US-0151438.
PR   01-SEP-1999; 99US-0151930.
PR   07-SEP-1999; 99US-0152363.
PR   07-SEP-1999; 99US-0153070.
PR   10-SEP-1999; 99US-0153758.
PR   13-SEP-1999; 99US-0154018.
PR   15-SEP-1999; 99US-0154039.
PR   16-SEP-1999; 99US-0154779.
PR   20-SEP-1999; 99US-0155139.
PR   22-SEP-1999; 99US-0155486.
PR   23-SEP-1999; 99US-0155659.
PR   24-SEP-1999; 99US-0156458.
PR   28-SEP-1999; 99US-0156596.
PR   29-SEP-1999; 99US-0157117.
PR   04-OCT-1999; 99US-0157753.
PR   05-OCT-1999; 99US-0157865.
PR   06-OCT-1999; 99US-0158029.
PR   07-OCT-1999; 99US-0158232.
PR   08-OCT-1999; 99US-0158369.
PR   12-OCT-1999; 99US-0159293.
PR   13-OCT-1999; 99US-0159294.
PR   13-OCT-1999; 99US-0159295.
PR   14-OCT-1999; 99US-0159329.
PR   14-OCT-1999; 99US-0159330.
PR   14-OCT-1999; 99US-0159931.
PR   14-OCT-1999; 99US-0159637.
PR   15-OCT-1999; 99US-0159638.
PR   18-OCT-1999; 99US-0159584.
PR   21-OCT-1999; 99US-0160741.
PR   21-OCT-1999; 99US-0160767.
PR   21-OCT-1999; 99US-0160768.
PR   21-OCT-1999; 99US-0160670.
PR   21-OCT-1999; 99US-0160814.
PR   21-OCT-1999; 99US-0160815.
PR   22-OCT-1999; 99US-0160980.
PR   22-OCT-1999; 99US-0160981.
PR   22-OCT-1999; 99US-0160989.
PR   25-OCT-1999; 99US-0161404.
PR   25-OCT-1999; 99US-0161405.
PR   25-OCT-1999; 99US-0161406.
PR   26-OCT-1999; 99US-0161359.
PR   26-OCT-1999; 99US-0161360.
PR   26-OCT-1999; 99US-0161361.
PR   28-OCT-1999; 99US-0161920.
PR   28-OCT-1999; 99US-0161992.
PR   28-OCT-1999; 99US-0161993.
PR   29-OCT-1999; 99US-0162142.

PR   25-FEB-1999; 99US-0121825.
PR   05-MAR-1999; 99US-0123180.
PR   09-MAR-1999; 99US-0123548.
PR   23-MAR-1999; 99US-0125788.
PR   25-MAR-1999; 99US-0126264.
PR   29-MAR-1999; 99US-0126785.
PR   01-APR-1999; 99US-0127462.
PR   06-APR-1999; 99US-0128234.
PR   08-APR-1999; 99US-0128714.
PR   16-APR-1999; 99US-0129845.
PR   19-APR-1999; 99US-0130077.
PR   21-APR-1999; 99US-0130449.
PR   23-APR-1999; 99US-0130510.
PR   28-APR-1999; 99US-0131449.
PR   30-APR-1999; 99US-0132048.
PR   04-MAY-1999; 99US-0132407.
PR   05-MAY-1999; 99US-0132484.
PR   05-MAY-1999; 99US-0132485.
PR   06-MAY-1999; 99US-0132486.
PR   06-MAY-1999; 99US-0132487.
PR   07-MAY-1999; 99US-0132863.
PR   11-MAY-1999; 99US-0134256.

```
PR   14-MAY-1999;   99US-0134218.        PR   23-JUL-1999;   99US-0145224.
PR   14-MAY-1999;   99US-0134219.        PR   26-JUL-1999;   99US-0145276.
PR   14-MAY-1999;   99US-0134221.        PR   27-JUL-1999;   99US-0145913.
PR   14-MAY-1999;   99US-0134370.        PR   27-JUL-1999;   99US-0145918.
PR   18-MAY-1999;   99US-0134768.        PR   27-JUL-1999;   99US-0145919.
PR   19-MAY-1999;   99US-0134941.        PR   28-JUL-1999;   99US-0145951.
PR   20-MAY-1999;   99US-0135124.        PR   02-AUG-1999;   99US-0146386.
PR   21-MAY-1999;   99US-0135353.        PR   02-AUG-1999;   99US-0146388.
PR   24-MAY-1999;   99US-0135629.        PR   02-AUG-1999;   99US-0146389.
PR   25-MAY-1999;   99US-0136021.        PR   02-AUG-1999;   99US-0146338.
PR   27-MAY-1999;   99US-0136392.        PR   03-AUG-1999;   99US-0147038.
PR   28-MAY-1999;   99US-0136782.        PR   04-AUG-1999;   99US-0147204.
PR   01-JUN-1999;   99US-0137222.        PR   04-AUG-1999;   99US-0147302.
PR   03-JUN-1999;   99US-0137528.        PR   05-AUG-1999;   99US-0147192.
PR   04-JUN-1999;   99US-0137502.        PR   05-AUG-1999;   99US-0147260.
PR   07-JUN-1999;   99US-0137724.        PR   06-AUG-1999;   99US-0147303.
PR   08-JUN-1999;   99US-0138094.        PR   06-AUG-1999;   99US-0147416.
PR   10-JUN-1999;   99US-0138540.        PR   09-AUG-1999;   99US-0147493.
PR   10-JUN-1999;   99US-0138847.        PR   09-AUG-1999;   99US-0147935.
PR   14-JUN-1999;   99US-0139119.        PR   10-AUG-1999;   99US-0147171.
PR   16-JUN-1999;   99US-0139452.        PR   11-AUG-1999;   99US-0148319.
PR   16-JUN-1999;   99US-0139453.        PR   12-AUG-1999;   99US-0148341.
PR   17-JUN-1999;   99US-0139492.        PR   13-AUG-1999;   99US-0148565.
PR   18-JUN-1999;   99US-0139454.        PR   13-AUG-1999;   99US-0148684.
PR   18-JUN-1999;   99US-0139455.        PR   16-AUG-1999;   99US-0149368.
PR   18-JUN-1999;   99US-0139456.        PR   17-AUG-1999;   99US-0149175.
PR   18-JUN-1999;   99US-0139457.        PR   18-AUG-1999;   99US-0149426.
PR   18-JUN-1999;   99US-0139458.        PR   20-AUG-1999;   99US-0149722.
PR   18-JUN-1999;   99US-0139459.        PR   20-AUG-1999;   99US-0149723.
PR   18-JUN-1999;   99US-0139460.        PR   23-AUG-1999;   99US-0149929.
PR   18-JUN-1999;   99US-0139461.        PR   23-AUG-1999;   99US-0149902.
PR   18-JUN-1999;   99US-0139462.        PR   25-AUG-1999;   99US-0150566.
PR   18-JUN-1999;   99US-0139463.        PR   26-AUG-1999;   99US-0150884.
PR   18-JUN-1999;   99US-0139750.        PR   27-AUG-1999;   99US-0151065.
PR   18-JUN-1999;   99US-0139763.        PR   27-AUG-1999;   99US-0151066.
PR   21-JUN-1999;   99US-0139817.        PR   27-AUG-1999;   99US-0151080.
PR   22-JUN-1999;   99US-0139899.        PR   30-AUG-1999;   99US-0151303.
PR   22-JUN-1999;   99US-0140353.        PR   31-AUG-1999;   99US-0151438.
PR   23-JUN-1999;   99US-0140354.        PR   01-SEP-1999;   99US-0151930.
PR   24-JUN-1999;   99US-0140695.        PR   07-SEP-1999;   99US-0151363.
PR   28-JUN-1999;   99US-0140823.        PR   10-SEP-1999;   99US-0153070.
PR   29-JUN-1999;   99US-0140991.        PR   13-SEP-1999;   99US-0153758.
PR   30-JUN-1999;   99US-0141287.        PR   15-SEP-1999;   99US-0154018.
PR   01-JUL-1999;   99US-0141842.        PR   16-SEP-1999;   99US-0154039.
PR   01-JUL-1999;   99US-0142154.        PR   20-SEP-1999;   99US-0154779.
PR   02-JUL-1999;   99US-0142055.        PR   22-SEP-1999;   99US-0155139.
PR   06-JUL-1999;   99US-0142390.        PR   23-SEP-1999;   99US-0155486.
PR   08-JUL-1999;   99US-0142803.        PR   24-SEP-1999;   99US-0155659.
PR   09-JUL-1999;   99US-0142920.        PR   29-SEP-1999;   99US-0156596.
PR   12-JUL-1999;   99US-0142977.        PR   04-OCT-1999;   99US-0157117.
PR   13-JUL-1999;   99US-0143542.        PR   05-OCT-1999;   99US-0157753.
PR   14-JUL-1999;   99US-0143624.        PR   06-OCT-1999;   99US-0157865.
PR   15-JUL-1999;   99US-0144005.        PR   07-OCT-1999;   99US-0158029.
PR   16-JUL-1999;   99US-0144085.        PR   08-OCT-1999;   99US-0158232.
PR   16-JUL-1999;   99US-0144086.        PR   12-OCT-1999;   99US-0158369.
PR   19-JUL-1999;   99US-0144325.        PR   13-OCT-1999;   99US-0159293.
PR   19-JUL-1999;   99US-0144331.        PR   13-OCT-1999;   99US-0159294.
PR   19-JUL-1999;   99US-0144332.        PR   14-OCT-1999;   99US-0159295.
PR   19-JUL-1999;   99US-0144333.        PR   14-OCT-1999;   99US-0159329.
PR   19-JUL-1999;   99US-0144334.        PR   14-OCT-1999;   99US-0159330.
PR   19-JUL-1999;   99US-0144335.        PR   14-OCT-1999;   99US-0159331.
PR   20-JUL-1999;   99US-0144352.        PR   14-OCT-1999;   99US-0159637.
PR   20-JUL-1999;   99US-0144632.        PR   18-OCT-1999;   99US-0159638.
PR   20-JUL-1999;   99US-0144884.        PR   21-OCT-1999;   99US-0159584.
PR   21-JUL-1999;   99US-0144814.        PR   21-OCT-1999;   99US-0160741.
PR   21-JUL-1999;   99US-0145086.        PR   21-OCT-1999;   99US-0160767.
PR   21-JUL-1999;   99US-0145088.        PR   21-OCT-1999;   99US-0160768.
PR   22-JUL-1999;   99US-0145085.        PR   21-OCT-1999;   99US-0160770.
PR   22-JUL-1999;   99US-0145087.        PR   22-OCT-1999;   99US-0160814.
PR   22-JUL-1999;   99US-0145089.        PR   22-OCT-1999;   99US-0160815.
PR   23-JUL-1999;   99US-0145145.        PR   22-OCT-1999;   99US-0160980.
PR   23-JUL-1999;   99US-0145192.        PR   22-OCT-1999;   99US-0160981.
PR   23-JUL-1999;   99US-0145218.
```

PR   22-OCT-1999; 99US-0160989.
PR   25-OCT-1999; 99US-0161404.
PR   25-OCT-1999; 99US-0161405.
PR   25-OCT-1999; 99US-0161406.
PR   26-OCT-1999; 99US-0161359.
PR   26-OCT-1999; 99US-0161360.
PR   26-OCT-1999; 99US-0161361.
PR   28-OCT-1999; 99US-0161920.
PR   28-OCT-1999; 99US-0161992.
PR   28-OCT-1999; 99US-0161993.
PR   29-OCT-1999; 99US-0162142.

Query Match
Best Local Similarity   31.6%; Score 608; DB 21; Length 330;
Matches 125; Conservative 63; Mismatches 80; Indels 58; Gaps 3;

```
Qy   88  EVVGGYLAHSLAVMTDAAHLLTDFASMLISLFSLWMSSRPATKIMNFGWQORAEILGALVS 147
         |||||         |||||||            ||||||||||||||||||||||||||||
Db    4  EVVGGIKANSLAILTDAAHLLSDVAAFAISLFSLWAAGWEATPRQTGFERIELGALVS 63

Qy  148  VLSIWVVTGVLYLVLAVERLISGDYEIDGGTMLITSGCAVAVNIIMGLTL-----HQSGHG 202
         |:||:|||:|:||||||||::|| |:||||||||||||||||||||||:        |||
Db   64  IQLIWLLTGILVVEAIIRIVTETSEVNGFLMFLVAAFGLVVNIIMAVLLGHDHGHSHGHG 123

Qy  203  HSHG------------------TTN-- 209
         ||:                   |:
Db  124  HGHGHDHNHNHSHGVTVTTHHHHDHEHGHSHGHGHEDKHHAHGDVTEQLLDKSKTQVAAKE 183

Qy  210  QQEENPSVRAAFIHVIGDFMQSMGVLVAAYILYFKPEYKYVDPICTFVFSILVLGTTLTI 269
         |||:|||||||||||||:|||:|||||||||:|||||  |:|:|||||||||:|||
Db  184  KRKRNINLQGAYLHVLGDSTIQSVGVMIGGAIIIWYNPEWKIVDLICTLAFSVIVLGTTINM 243

Qy  270  LRDVILVLMEGTPKGVDFTAVRDLLLSVEGVEALHSLHIWALTVAQPVLSVHIAIAQNTD 329
         |:|:|| ||||::||||||||||||||:|||||||::||||||  ||::| |:||:|||
Db  244  IRNLIEVLMESTPREIDATKLEKGLLEMEEVVAVHELHIWAITVGKVLLACHVNIRPEAD 303

Qy  330  AQAVLKTASSRLQGKFHFHHVTIQIE 355
         |||||||||||||||||||||||||
Db  304  ADMVLNKVIDYIRREYNISHVTIQIE 329
```

RESULT 23
ABB65511
ID   ABB65511 standard; Protein; 452 AA.
XX
AC   ABB65511;
XX
DT   26-MAR-2002 (first entry)
XX
DE   Drosophila melanogaster polypeptide SEQ ID NO 23325.
XX
KW   Drosophila; developmental biology; cell signalling; insecticide;
KW   pharmaceutical.
XX
OS   Drosophila melanogaster.
XX
PN   WO200171042-A2.
XX
PD   27-SEP-2001.
XX
PI   Venter JC, Adams M, Li PWD, Myers EW;
XX
PA   (PEKE) PE CORP NY.
XX
PR   23-MAR-2000; 2000US-191637P.
PR   11-JUL-2000; 2000US-0614150.
XX
PF   23-MAR-2001: 2001WO-US09231.
XX
DR   WPI; 2001-656860/75.
DR   N-PSDB; ABL09614.
XX
PT   New isolated nucleic acid detection reagent for detecting 1000 or more

genes from Drosophila and for elucidating cell signalling and cell-cell
interactions -
PT
XX
PS   Disclosure; SEQ ID NO 23325; 2lpp + Sequence Listing; English.
XX
CC   The invention relates to an isolated nucleic acid detection reagent
CC   capable of detecting 1000 or more genes from Drosophila. The invention is
CC   useful in developmental biology and in elucidating cell signaling and
CC   cell-cell interactions in higher eukaryotes for the development of
CC   insecticides, therapeutics and pharmaceutical drugs. The invention
CC   discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA
CC   sequences (ABL01840-ABL16175) and the encoded proteins
CC   (ABB57737-ABB72072).
CC   The sequence data for this patent did not form part of the printed
CC   specification, but was obtained in electronic format directly from WIPO
CC   at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ   Sequence   452 AA;

Query Match
Best Local Similarity   29.4%; Score 565; DB 22; Length 452;
Matches 118; Conservative 66; Mismatches 88; Indels 80; Gaps 4;

```
Qy   88  EVVGGYLAHSLAVMTDAAHLLTDFASMLISLFSLWMSSRPATKIMNFGWQORAEILGALVS 147
         :||||:   |:|:||||||| :|  |::  ||: :| | | | ||        |:|||:
Db  101  EFLGGYVAGSLAIMTDAAHLASDCISFVIGLVAIWIGGRPPDDERMSFGYKRFEVIGALAS 160

Qy  148  VLSIWVVTGVLYLVLAVERLISGDYEIDGGTMLITSGCAVAVNIIMGLTLHQS------GH 201
         :| || || :|:| |||||: ::|    | ||:|| |   |:| |::|  |:       ||
Db  161  ILGIWFVTTLLVVVAIERIFSQDFELANDMMLISGIGIVINIVMMFVLHGSWFVNGNGH 220

Qy  202  GHSH------------------ 205
         ||||
Db  221  GHSHSHSHSHSHSHGNGHEPNDSLSQTRSNSNNFLTTIGSQSASTADEEDSIRKEINSNEH 280

Qy  206  ------GTTN----QQEENPSVRAAFIHVIGDFMQSMGVLVAAYILYFKPEYK 248
             ||:|    |||:|||||||||||||:||||:|||||||||:||||
Db  281  KIVITNGKKPTLTGTSNMELAHEDKNLNRAAMIHVIGDLVQSIGVFLAAVLIKVCPGAK 340

Qy  249  YVDPICTFVFSILVLGTTLTILRDVILVLMEGTPKGVDFTAVRDLLLSVEGVEALHSLHI 308
         |.||:|| :|||:| : | |:|| : |:  |  ||:|    : |   |:|:  |:| |:
Db  341  YADPLCTLIFSIIVIMFTLRLRESLGIIVNAVPQNLNMRTLHLELGSIEGVRSLHHLNV 400

Qy  309  WALTVAQPVLSVHIAIAQNTDAQAVLKTASSRLQG-KFHFHTVTIQIEDYSE 359
         ||:| :| ||:|| | |::  : || ::| :   | :| |:||:|:| | :|
Db  401  WQQTSQQRVLMVHLVTDSRADGNEVLQAATALVSSPRYNIKHSTIQLEPTTE 452
```

RESULT 24
AAG31822
ID   AAG31822 standard; Protein; 375 AA.
XX
AC   AAG31822;
XX
DT   17-OCT-2000 (first entry)
XX
DE   Arabidopsis thaliana protein fragment SEQ ID NO: 38280.
XX
KW   Protein identification; signal transduction pathway; metabolic
KW   pathway;
KW   hybridisation assay; genetic mapping; gene expression control; promoter;
KW   termination sequence.
XX
OS   Arabidopsis thaliana.
XX
PN   EP1033405-A2.
XX
PD   06-SEP-2000.
XX
PF   25-FEB-2000; 2000EP-0301439.
XX
PR   25-FEB-1999; 99US-0121825.
PR   05-MAR-1999; 99US-0123180.
PR   09-MAR-1999; 99US-0123548.

```
PR  23-MAR-1999;  99US-0125788.
PR  25-MAR-1999;  99US-0126264.
PR  29-MAR-1999;  99US-0126785.
PR  01-APR-1999;  99US-0127462.
PR  06-APR-1999;  99US-0128234.
PR  08-APR-1999;  99US-0128714.
PR  16-APR-1999;  99US-0129845.
PR  19-APR-1999;  99US-0130077.
PR  21-APR-1999;  99US-0130449.
PR  23-APR-1999;  99US-0130510.
PR  28-APR-1999;  99US-0310891.
PR  30-APR-1999;  99US-0311449.
PR  30-APR-1999;  99US-0132048.
PR  04-MAY-1999;  99US-0132407.
PR  05-MAY-1999;  99US-0132484.
PR  06-MAY-1999;  99US-0132485.
PR  06-MAY-1999;  99US-0132486.
PR  07-MAY-1999;  99US-0132487.
PR  11-MAY-1999;  99US-0132863.
PR  14-MAY-1999;  99US-0134256.
PR  14-MAY-1999;  99US-0134218.
PR  14-MAY-1999;  99US-0134219.
PR  25-MAY-1999;  99US-0136021.
PR  27-MAY-1999;  99US-0136392.
PR  28-MAY-1999;  99US-0136782.
PR  01-JUN-1999;  99US-0136540.
PR  01-JUN-1999;  99US-0137222.
PR  03-JUN-1999;  99US-0137528.
PR  04-JUN-1999;  99US-0137502.
PR  07-JUN-1999;  99US-0137724.
PR  08-JUN-1999;  99US-0138094.
PR  10-JUN-1999;  99US-0138540.
PR  10-JUN-1999;  99US-0138847.
PR  14-JUN-1999;  99US-0139119.
PR  16-JUN-1999;  99US-0139452.
PR  16-JUN-1999;  99US-0139453.
PR  17-JUN-1999;  99US-0139492.
PR  18-JUN-1999;  99US-0139454.
PR  18-JUN-1999;  99US-0139455.
PR  18-JUN-1999;  99US-0139456.
PR  18-JUN-1999;  99US-0139457.
PR  18-JUN-1999;  99US-0139458.
PR  18-JUN-1999;  99US-0139459.
PR  18-JUN-1999;  99US-0139460.
PR  18-JUN-1999;  99US-0139461.
PR  18-JUN-1999;  99US-0139462.
PR  18-JUN-1999;  99US-0139463.
PR  18-JUN-1999;  99US-0139750.
PR  21-JUN-1999;  99US-0139763.
PR  22-JUN-1999;  99US-0139817.
PR  23-JUN-1999;  99US-0139899.
PR  23-JUN-1999;  99US-0140353.
PR  24-JUN-1999;  99US-0140354.
PR  28-JUN-1999;  99US-0140695.
PR  29-JUN-1999;  99US-0140823.
PR  30-JUN-1999;  99US-0140991.
PR  01-JUL-1999;  99US-0141287.
PR  01-JUL-1999;  99US-0141842.
PR  02-JUL-1999;  99US-0142154.
PR  06-JUL-1999;  99US-0142055.
PR  08-JUL-1999;  99US-0142390.
PR  09-JUL-1999;  99US-0142803.
PR  09-JUL-1999;  99US-0142920.
PR  12-JUL-1999;  99US-0142977.
PR  13-JUL-1999;  99US-0143542.
PR  14-JUL-1999;  99US-0143624.
PR  15-JUL-1999;  99US-0144005.

PR  16-JUL-1999;  99US-0144085.
PR  16-JUL-1999;  99US-0144086.
PR  19-JUL-1999;  99US-0144325.
PR  19-JUL-1999;  99US-0144331.
PR  19-JUL-1999;  99US-0144332.
PR  19-JUL-1999;  99US-0144333.
PR  19-JUL-1999;  99US-0144334.
PR  19-JUL-1999;  99US-0144335.
PR  20-JUL-1999;  99US-0143352.
PR  20-JUL-1999;  99US-0144632.
PR  20-JUL-1999;  99US-0144884.
PR  21-JUL-1999;  99US-0144814.
PR  21-JUL-1999;  99US-0145086.
PR  22-JUL-1999;  99US-0145085.
PR  22-JUL-1999;  99US-0145087.
PR  22-JUL-1999;  99US-0145089.
PR  22-JUL-1999;  99US-0145192.
PR  23-JUL-1999;  99US-0145145.
PR  23-JUL-1999;  99US-0145218.
PR  23-JUL-1999;  99US-0145224.
PR  26-JUL-1999;  99US-0145276.
PR  27-JUL-1999;  99US-0145513.
PR  27-JUL-1999;  99US-0145918.
PR  28-JUL-1999;  99US-0145919.
PR  28-JUL-1999;  99US-0145951.
PR  02-AUG-1999;  99US-0145386.
PR  02-AUG-1999;  99US-0146388.
PR  02-AUG-1999;  99US-0146389.
PR  03-AUG-1999;  99US-0147038.
PR  04-AUG-1999;  99US-0147204.
PR  05-AUG-1999;  99US-0147202.
PR  05-AUG-1999;  99US-0147260.
PR  06-AUG-1999;  99US-0147303.
PR  06-AUG-1999;  99US-0147416.
PR  09-AUG-1999;  99US-0147493.
PR  09-AUG-1999;  99US-0147935.
PR  10-AUG-1999;  99US-0148171.
PR  11-AUG-1999;  99US-0148319.
PR  12-AUG-1999;  99US-0148341.
PR  13-AUG-1999;  99US-0148565.
PR  13-AUG-1999;  99US-0148684.
PR  16-AUG-1999;  99US-0149175.
PR  17-AUG-1999;  99US-0149426.
PR  18-AUG-1999;  99US-0149722.
PR  20-AUG-1999;  99US-0149723.
PR  20-AUG-1999;  99US-0149929.
PR  23-AUG-1999;  99US-0149902.
PR  23-AUG-1999;  99US-0150884.
PR  25-AUG-1999;  99US-0150566.
PR  26-AUG-1999;  99US-0151065.
PR  27-AUG-1999;  99US-0151066.
PR  27-AUG-1999;  99US-0151080.
PR  30-AUG-1999;  99US-0151303.
PR  31-AUG-1999;  99US-0151438.
PR  01-SEP-1999;  99US-0151930.
PR  07-SEP-1999;  99US-0152263.
PR  10-SEP-1999;  99US-0153070.
PR  13-SEP-1999;  99US-0153758.
PR  15-SEP-1999;  99US-0154018.
PR  16-SEP-1999;  99US-0154039.
PR  20-SEP-1999;  99US-0154779.
PR  22-SEP-1999;  99US-0155139.
PR  23-SEP-1999;  99US-0155486.
PR  24-SEP-1999;  99US-0155559.
PR  28-SEP-1999;  99US-0156458.
PR  29-SEP-1999;  99US-0155596.
PR  04-OCT-1999;  99US-0157117.
PR  05-OCT-1999;  99US-0157753.
PR  06-OCT-1999;  99US-0157865.
```

```
RESULT 25
AAG31821
ID   AAG31821 standard; Protein; 385 AA.
AC   AAG31821;
XX
DT   17-OCT-2000 (first entry)
XX
```

```
DE   Arabidopsis thaliana protein fragment SEQ ID NO: 38279.
XX
KW   Protein identification; signal transduction pathway; metabolic pathway;
KW   hybridisation assay; genetic mapping; gene expression control; promoter;
KW   termination sequence.
XX
OS   Arabidopsis thaliana.
XX
PN   EP1033405-A2.
XX
PD   06-SEP-2000.
XX
PF   25-FEB-2000; 2000EP-0301439.
```

```
PR   07-OCT-1999; 99US-0158029.          PR   25-FEB-1999; 99US-0121825.
PR   08-OCT-1999; 99US-0158232.          PR   05-MAR-1999; 99US-0123180.
PR   12-OCT-1999; 99US-0158369.          PR   09-MAR-1999; 99US-0123548.
PR   13-OCT-1999; 99US-0159293.          PR   23-MAR-1999; 99US-0125788.
PR   13-OCT-1999; 99US-0159294.          PR   25-MAR-1999; 99US-0126264.
PR   13-OCT-1999; 99US-0159295.          PR   29-MAR-1999; 99US-0126785.
PR   14-OCT-1999; 99US-0159329.          PR   01-APR-1999; 99US-0127462.
PR   14-OCT-1999; 99US-0159330.          PR   06-APR-1999; 99US-0128234.
PR   14-OCT-1999; 99US-0159331.          PR   08-APR-1999; 99US-0128714.
PR   14-OCT-1999; 99US-0159637.          PR   16-APR-1999; 99US-0129845.
PR   14-OCT-1999; 99US-0159638.          PR   19-APR-1999; 99US-0130077.
PR   18-OCT-1999; 99US-0159584.          PR   21-APR-1999; 99US-0130449.
PR   18-OCT-1999; 99US-0159585.          PR   23-APR-1999; 99US-0130510.
PR   21-OCT-1999; 99US-0160741.          PR   28-APR-1999; 99US-0130891.
PR   21-OCT-1999; 99US-0160767.          PR   30-APR-1999; 99US-0131449.
PR   21-OCT-1999; 99US-0160768.          PR   30-APR-1999; 99US-0132048.
PR   21-OCT-1999; 99US-0160770.          PR   04-MAY-1999; 99US-0132407.
PR   21-OCT-1999; 99US-0160814.          PR   04-MAY-1999; 99US-0132484.
PR   21-OCT-1999; 99US-0160815.          PR   05-MAY-1999; 99US-0132485.
PR   22-OCT-1999; 99US-0160980.          PR   06-MAY-1999; 99US-0132486.
PR   22-OCT-1999; 99US-0160981.          PR   06-MAY-1999; 99US-0132487.
PR   22-OCT-1999; 99US-0160989.          PR   07-MAY-1999; 99US-0132863.
PR   25-OCT-1999; 99US-0161404.          PR   11-MAY-1999; 99US-0134256.
PR   25-OCT-1999; 99US-0161405.          PR   14-MAY-1999; 99US-0134218.
PR   25-OCT-1999; 99US-0161406.          PR   14-MAY-1999; 99US-0134219.
PR   26-OCT-1999; 99US-0161359.          PR   14-MAY-1999; 99US-0134221.
PR   26-OCT-1999; 99US-0161360.          PR   14-MAY-1999; 99US-0134370.
PR   26-OCT-1999; 99US-0161361.          PR   19-MAY-1999; 99US-0134748.
PR   28-OCT-1999; 99US-0161920.          PR   20-MAY-1999; 99US-0134941.
PR   28-OCT-1999; 99US-0161992.          PR   21-MAY-1999; 99US-0135124.
PR   28-OCT-1999; 99US-0161993.          PR   21-MAY-1999; 99US-0135353.
PR   29-OCT-1999; 99US-0162142.          PR   24-MAY-1999; 99US-0135629.
                                          PR   25-MAY-1999; 99US-0136021.
                                          PR   27-MAY-1999; 99US-0136392.
                                          PR   28-MAY-1999; 99US-0136782.
                                          PR   01-JUN-1999; 99US-0137222.
                                          PR   03-JUN-1999; 99US-0137528.
                                          PR   04-JUN-1999; 99US-0137502.
                                          PR   07-JUN-1999; 99US-0137724.
                                          PR   08-JUN-1999; 99US-0138094.
                                          PR   10-JUN-1999; 99US-0138540.
                                          PR   10-JUN-1999; 99US-0138847.
                                          PR   14-JUN-1999; 99US-0139119.
                                          PR   16-JUN-1999; 99US-0139452.
                                          PR   16-JUN-1999; 99US-0139453.
                                          PR   17-JUN-1999; 99US-0139492.
                                          PR   18-JUN-1999; 99US-0139454.
                                          PR   18-JUN-1999; 99US-0139455.
                                          PR   18-JUN-1999; 99US-0139456.
                                          PR   18-JUN-1999; 99US-0139457.
                                          PR   18-JUN-1999; 99US-0139458.
                                          PR   18-JUN-1999; 99US-0139459.
                                          PR   18-JUN-1999; 99US-0139460.
                                          PR   18-JUN-1999; 99US-0139461.
                                          PR   18-JUN-1999; 99US-0139462.
                                          PR   18-JUN-1999; 99US-0139463.
                                          PR   18-JUN-1999; 99US-0139750.
                                          PR   18-JUN-1999; 99US-0139763.
                                          PR   21-JUN-1999; 99US-0139817.
```

```
Query Match             26.0%;  Score 500.5;  DB 21;  Length 375;
Best Local Similarity   29.8%;  Pred. No. 5.6e-45;
Matches  114;  Conservative   81;  Mismatches 130;  Indels  57;  Gaps   5;

Qy    11 DARPAIRSYTGSLWQEGAGWIPLPRPGLDLQAIELAAQSNHHCHAQKGPDSHCDPKKGKA 70
            ::| | |            |        |:||
Db    12 DDEEEMESSPSPSKTEENLGVVPL------------SCAFTRQEHCVSE------TKEREES 54

Qy    71 QRQLVASAICLLFMIGEVVGGYLAHSLAVMTDAAHLLTDFASMLISLFSLWMSSRPATK 130
          ::|| |::| ::|: |:| || | ||||||||||| :|::::::| ||
Db    55 TRRLSSLIFLYLIVMSVQIVGGFKANSLAVMTDAAHLLSDVAGLCVSLLAIKVSSWEANP 114

Qy   131 TMNFGWQRAEILGALVSVSLSIWVVTGVLVYLAVERLISGDYEIDGGTMLITSGCAVAVNI 190
           ::| |:   |  |::|  |: |:|:|::: | | :||:| |::    ||   | |:
Db   115 RNSFGFKRLEVLAAFLSVQLIWLVSGVIIHEAIQRLLSRSREVNGEIMFGISAFGFFMNL 174

Qy   191 IMGLTLHQSGHGHSH------------GTTNQEENP------------ 215
          :|| |  |  |  |||               ::  ||
Db   175 VMVLWL---GHNHSHHHDHHHHHHHHHHHKHQHQHHHKEVVAEEEEEEMNPLKGEKSSSKEM 231

Qy   216 --SVRAAFIHVIGDFMQSMGVLVAAYILYFKPEYKXVDPICTFVFSILVLGTTLTILRDV 273
            ::|:|||:|   |::|||||:|||||| | |||::|:||:||:|  | ||| ||::| |
Db   232 NINIQGAYLHAMADMIQSLGVMIGGIIWVKPKWLVDLICTLVFSAFALAATLPLKNI 291

Qy   274 ILVLMEGTPKGVDFTAVRDLLLSVEGGVEALHSLHIWALTVAQPVLSVHIAIAQNTDAQAV 333
          |::| | ||:  | |: ||::||:::  |  |:| |:|: || | |:| |||: :  |||:
Db   292 FGILMERVPRDMDIEKLERGLKRIDGVKIVYDLHVWEITVGRIVLSCHILPEPGASPKEI 351

Qy   334 LKTASSRLQGSKFHFHTVTIQIE 355
          :||: |  |  | ||||:|:|:|
Db   352 ITGVRNFCRKSYGIYHATVQVE 373
```

```
PR  22-JUN-1999;  99US-0139899.
PR  23-JUN-1999;  99US-0140353.
PR  23-JUN-1999;  99US-0140354.
PR  24-JUN-1999;  99US-0140695.
PR  28-JUN-1999;  99US-0140823.
PR  29-JUN-1999;  99US-0140991.
PR  30-JUN-1999;  99US-0141287.
PR  01-JUL-1999;  99US-0141842.
PR  01-JUL-1999;  99US-0142154.
PR  02-JUL-1999;  99US-0142055.
PR  06-JUL-1999;  99US-0142390.
PR  08-JUL-1999;  99US-0142803.
PR  09-JUL-1999;  99US-0142920.
PR  12-JUL-1999;  99US-0142977.
PR  13-JUL-1999;  99US-0143542.
PR  14-JUL-1999;  99US-0143624.
PR  15-JUL-1999;  99US-0144005.
PR  16-JUL-1999;  99US-0144085.
PR  16-JUL-1999;  99US-0144086.
PR  19-JUL-1999;  99US-0144325.
PR  19-JUL-1999;  99US-0144331.
PR  19-JUL-1999;  99US-0144332.
PR  19-JUL-1999;  99US-0144333.
PR  19-JUL-1999;  99US-0144334.
PR  19-JUL-1999;  99US-0144335.
PR  20-JUL-1999;  99US-0144352.
PR  20-JUL-1999;  99US-0144632.
PR  21-JUL-1999;  99US-0144814.
PR  21-JUL-1999;  99US-0144884.
PR  21-JUL-1999;  99US-0145086.
PR  21-JUL-1999;  99US-0145088.
PR  22-JUL-1999;  99US-0145085.
PR  22-JUL-1999;  99US-0145087.
PR  22-JUL-1999;  99US-0145089.
PR  22-JUL-1999;  99US-0145192.
PR  23-JUL-1999;  99US-0145145.
PR  23-JUL-1999;  99US-0145218.
PR  23-JUL-1999;  99US-0145224.
PR  26-JUL-1999;  99US-0145276.
PR  27-JUL-1999;  99US-0145913.
PR  27-JUL-1999;  99US-0145918.
PR  27-JUL-1999;  99US-0145919.
PR  28-JUL-1999;  99US-0145951.
PR  02-AUG-1999;  99US-0146386.
PR  02-AUG-1999;  99US-0146388.
PR  02-AUG-1999;  99US-0146389.
PR  03-AUG-1999;  99US-0147038.
PR  04-AUG-1999;  99US-0147204.
PR  04-AUG-1999;  99US-0147302.
PR  05-AUG-1999;  99US-0147192.
PR  05-AUG-1999;  99US-0147260.
PR  06-AUG-1999;  99US-0147303.
PR  06-AUG-1999;  99US-0147416.
PR  09-AUG-1999;  99US-0147493.
PR  09-AUG-1999;  99US-0147935.
PR  10-AUG-1999;  99US-0148171.
PR  11-AUG-1999;  99US-0148319.
PR  12-AUG-1999;  99US-0148341.
PR  13-AUG-1999;  99US-0148565.
PR  13-AUG-1999;  99US-0148684.
PR  16-AUG-1999;  99US-0149368.
PR  17-AUG-1999;  99US-0149175.
PR  18-AUG-1999;  99US-0149426.
PR  20-AUG-1999;  99US-0149722.
PR  20-AUG-1999;  99US-0149723.
PR  20-AUG-1999;  99US-0149929.
PR  23-AUG-1999;  99US-0149902.
PR  23-AUG-1999;  99US-0149930.
PR  25-AUG-1999;  99US-0150566.
PR  26-AUG-1999;  99US-0150884.
PR  27-AUG-1999;  99US-0151065.
PR  27-AUG-1999;  99US-0151066.
PR  27-AUG-1999;  99US-0151080.
```

```
PR  30-AUG-1999;  99US-0151103.
PR  31-AUG-1999;  99US-0151438.
PR  01-SEP-1999;  99US-0151930.
PR  07-SEP-1999;  99US-0122363.
PR  10-SEP-1999;  99US-0153070.
PR  13-SEP-1999;  99US-0153758.
PR  15-SEP-1999;  99US-0154018.
PR  16-SEP-1999;  99US-0154039.
PR  20-SEP-1999;  99US-0154779.
PR  22-SEP-1999;  99US-0155139.
PR  23-SEP-1999;  99US-0155486.
PR  24-SEP-1999;  99US-0155659.
PR  28-SEP-1999;  99US-0156596.
PR  29-SEP-1999;  99US-0156458.
PR  04-OCT-1999;  99US-0157117.
PR  05-OCT-1999;  99US-0157753.
PR  06-OCT-1999;  99US-0157865.
PR  07-OCT-1999;  99US-0158029.
PR  08-OCT-1999;  99US-0158232.
PR  12-OCT-1999;  99US-0158369.
PR  13-OCT-1999;  99US-0159293.
PR  13-OCT-1999;  99US-0159294.
PR  13-OCT-1999;  99US-0159295.
PR  14-OCT-1999;  99US-0159329.
PR  14-OCT-1999;  99US-0159330.
PR  14-OCT-1999;  99US-0159331.
PR  14-OCT-1999;  99US-0159637.
PR  18-OCT-1999;  99US-0159638.
PR  18-OCT-1999;  99US-0159584.
PR  21-OCT-1999;  99US-0160741.
PR  21-OCT-1999;  99US-0160767.
PR  21-OCT-1999;  99US-0160768.
PR  21-OCT-1999;  99US-0160770.
PR  21-OCT-1999;  99US-0160814.
PR  21-OCT-1999;  99US-0160815.
PR  22-OCT-1999;  99US-0160980.
PR  22-OCT-1999;  99US-0160981.
PR  22-OCT-1999;  99US-0160989.
PR  25-OCT-1999;  99US-0161404.
PR  25-OCT-1999;  99US-0161405.
PR  25-OCT-1999;  99US-0161406.
PR  26-OCT-1999;  99US-0161359.
PR  26-OCT-1999;  99US-0161360.
PR  26-OCT-1999;  99US-0161361.
PR  28-OCT-1999;  99US-0161920.
PR  28-OCT-1999;  99US-0161992.
PR  28-OCT-1999;  99US-0161993.
PR  29-OCT-1999;  99US-0162142.
```

```
Query Match         26.0%;    Score 500.5;   DB 21;   Length 385;
Best Local Similarity  29.8%;   Pred. No. 5.8e-45;
Matches  114;  Conservative  81;  Mismatches 130;  Indels  57;  Gaps  5;

Qy  11 DARPAIRSYTGSLWQEGAGWIPLPRPGLDLQAIELAAQSNHHCHAQKGPDSHCDPKKGKA 70
           :  ||  :  | ||  |  :| ||||||||||: | :  :   ::   :
Db  22 DDEEMESPSPSKTEENLGVVPL---------------SCAFTRQEHCVSE------TKEREES 64

Qy  71 QRQLYVASAICLLFMIGEVVGGYLAHSLAVMTDAAHLLTDFASMLISLFSLWMSSRPATK 130
           |::||||  |:||| | ||||: |:||||||||||||: ::  :| :|::|  |||
Db  65 TRRLSSLIFLYLIVMSVQIVGGFKANSLAVMTDAAHLLSDVAGLCVSLLAIKVSSWEANP 124

Qy 131 TMNFGWQRAEILCALVSVLSIWVVTGVLVYLAVERLISGDYEIDGGTMLITSGCAVAVNI 190
           :||:|  |   |   |:  |  || : | ||:::| ||:|| |  ::  ||  |  :|
Db 125 RNSFGFKRLEVLAAFLSVQLIWLVSGVIIHEAIQRLLSRSREVNGEIMFGISAFGFFMNL 184

Qy 191 IMGLTLHQSGHGHSH------------------GTTNQEENP---------------215
           :|| |  ||| |
Db 185 VMVLWL---GHNHSHHHHDHHHHHHHNHKHQHHHKEVVAEEEEEEMPLKGEKSSSKEM 241

Qy 216 --SVRAAFTHVIGDFMQSMGVLVAAYILYFKPEYKYVDPICTFVFSILVLGTTLTILRDV 273
           :  | :|| |:  || : | || :| ||  ||:::
Db 242 NINQGAYLHAMADMIQSLGVMIGGGIIWVKPKWVLVDLICTLVFSAFALAATLPILKNI 301
```

```
Qy    274 ILVLMEGTPKGVDFTAVRDLLLSVEGVEALHSLHIWALTVAQPVLSSVHIAIAQNTDAQAV 333
          ||||  |:|||  |: |||  :||:|| ||:||| ::|| ||| ||
Db    302 FGILMERVPRDMDIEKLERGLKRIDGVKIYDLHVWEITVGRIVLSCHILPEPGASPKEI 361

Qy    334 LKTASSRLQGKFHFHTVTIQUE 355
          | |:|:|
Db    362 ITGVRNFCRKSYGIYHATVQVE 383

RESULT 26
AAG31823
ID   AAG31823 standard; Protein; 359 AA.
XX
AC   AAG31823;
XX
DT   17-OCT-2000   (first entry)
XX
DE   Arabidopsis thaliana protein fragment SEQ ID NO: 38281.
XX
KW   Protein identification; signal transduction pathway; metabolic pathway;
KW   hybridisation assay; genetic mapping; gene expression control; promoter;
KW   termination sequence.
XX
OS   Arabidopsis thaliana.
XX
PN   EP1033405-A2.
XX
PD   06-SEP-2000.
XX
PF   25-FEB-2000; 2000EP-0301439.
PR   25-FEB-1999;  99US-0121825.
PR   05-MAR-1999;  99US-0123180.
PR   09-MAR-1999;  99US-0123548.
PR   23-MAR-1999;  99US-0125788.
PR   25-MAR-1999;  99US-0126264.
PR   29-MAR-1999;  99US-0126785.
PR   01-APR-1999;  99US-0127462.
PR   06-APR-1999;  99US-0128234.
PR   08-APR-1999;  99US-0128714.
PR   16-APR-1999;  99US-0129845.
PR   19-APR-1999;  99US-0130077.
PR   21-APR-1999;  99US-0130449.
PR   23-APR-1999;  99US-0130510.
PR   28-APR-1999;  99US-0130891.
PR   30-APR-1999;  99US-0131449.
PR   30-APR-1999;  99US-0132048.
PR   04-MAY-1999;  99US-0132407.
PR   05-MAY-1999;  99US-0132484.
PR   05-MAY-1999;  99US-0132485.
PR   06-MAY-1999;  99US-0132486.
PR   07-MAY-1999;  99US-0132487.
PR   11-MAY-1999;  99US-0132863.
PR   14-MAY-1999;  99US-0134256.
PR   14-MAY-1999;  99US-0134218.
PR   14-MAY-1999;  99US-0134219.
PR   12-MAY-1999;  99US-0134221.
PR   14-MAY-1999;  99US-0134370.
PR   18-MAY-1999;  99US-0134768.
PR   19-MAY-1999;  99US-0134941.
PR   20-MAY-1999;  99US-0135124.
PR   21-MAY-1999;  99US-0135533.
PR   24-MAY-1999;  99US-0135629.
PR   25-MAY-1999;  99US-0136021.
PR   27-MAY-1999;  99US-0136392.
PR   28-MAY-1999;  99US-0136782.
PR   01-JUN-1999;  99US-0137222.
PR   03-JUN-1999;  99US-0137528.
PR   04-JUN-1999;  99US-0137502.
PR   07-JUN-1999;  99US-0137724.
PR   08-JUN-1999;  99US-0138004.
PR   10-JUN-1999;  99US-0138540.
PR   10-JUN-1999;  99US-0138847.
PR   14-JUN-1999;  99US-0139119.
PR   16-JUN-1999;  99US-0139452.
PR   16-JUN-1999;  99US-0139453.
PR   17-JUN-1999;  99US-0139492.
PR   18-JUN-1999;  99US-0139454.
PR   18-JUN-1999;  99US-0139455.
PR   18-JUN-1999;  99US-0139456.
PR   18-JUN-1999;  99US-0139457.
PR   18-JUN-1999;  99US-0139458.
PR   18-JUN-1999;  99US-0139459.
PR   18-JUN-1999;  99US-0139460.
PR   16-JUN-1999;  99US-0139461.
PR   18-JUN-1999;  99US-0139462.
PR   18-JUN-1999;  99US-0139463.
PR   21-JUN-1999;  99US-0139750.
PR   22-JUN-1999;  99US-0139763.
PR   21-JUN-1999;  99US-0139817.
PR   22-JUN-1999;  99US-0139899.
PR   23-JUN-1999;  99US-0140353.
PR   23-JUN-1999;  99US-0140354.
PR   24-JUN-1999;  99US-0140695.
PR   28-JUN-1999;  99US-0140823.
PR   29-JUN-1999;  99US-0140991.
PR   30-JUN-1999;  99US-0141287.
PR   01-JUL-1999;  99US-0141842.
PR   01-JUL-1999;  99US-0142154.
PR   02-JUL-1999;  99US-0142055.
PR   02-JUL-1999;  99US-0142390.
PR   06-JUL-1999;  99US-0142803.
PR   08-JUL-1999;  99US-0142920.
PR   09-JUL-1999;  99US-0142977.
PR   12-JUL-1999;  99US-0143542.
PR   13-JUL-1999;  99US-0143624.
PR   14-JUL-1999;  99US-0144005.
PR   15-JUL-1999;  99US-0144085.
PR   16-JUL-1999;  99US-0144086.
PR   16-JUL-1999;  99US-0144325.
PR   19-JUL-1999;  99US-0144326.
PR   19-JUL-1999;  99US-0144331.
PR   19-JUL-1999;  99US-0144332.
PR   19-JUL-1999;  99US-0144333.
PR   19-JUL-1999;  99US-0144334.
PR   19-JUL-1999;  99US-0144335.
PR   20-JUL-1999;  99US-0144352.
PR   20-JUL-1999;  99US-0144632.
PR   20-JUL-1999;  99US-0144884.
PR   21-JUL-1999;  99US-0144814.
PR   21-JUL-1999;  99US-0145086.
PR   22-JUL-1999;  99US-0145086.
PR   22-JUL-1999;  99US-0145088.
PR   22-JUL-1999;  99US-0145085.
PR   22-JUL-1999;  99US-0145087.
PR   23-JUL-1999;  99US-0145089.
PR   23-JUL-1999;  99US-0145145.
PR   23-JUL-1999;  99US-0145192.
PR   23-JUL-1999;  99US-0145218.
PR   26-JUL-1999;  99US-0145224.
PR   27-JUL-1999;  99US-0145276.
PR   27-JUL-1999;  99US-0145913.
PR   04-AUG-1999;  99US-0145918.
PR   04-AUG-1999;  99US-0145919.
PR   27-JUL-1999;  99US-0145951.
PR   28-JUL-1999;  99US-0146386.
PR   02-AUG-1999;  99US-0146388.
PR   02-AUG-1999;  99US-0146389.
PR   03-AUG-1999;  99US-0147038.
PR   03-AUG-1999;  99US-0147204.
PR   04-AUG-1999;  99US-0147302.
PR   05-AUG-1999;  99US-0147192.
PR   05-AUG-1999;  99US-0147260.
PR   05-AUG-1999;  99US-0147303.
PR   06-AUG-1999;  99US-0147416.
PR   06-AUG-1999;  99US-0147493.
PR   09-AUG-1999;  99US-0147935.
PR   10-AUG-1999;  99US-0148171.
```

```
PR   11-AUG-1999;  99US-0148319.
PR   12-AUG-1999;  99US-0148341.
PR   13-AUG-1999;  99US-0148565.
PR   13-AUG-1999;  99US-0148684.
PR   16-AUG-1999;  99US-0149368.
PR   17-AUG-1999;  99US-0149175.
PR   18-AUG-1999;  99US-0149426.
PR   18-AUG-1999;  99US-0149722.
PR   20-AUG-1999;  99US-0149723.
PR   20-AUG-1999;  99US-0149929.
PR   20-AUG-1999;  99US-0149902.
PR   23-AUG-1999;  99US-0149930.
PR   23-AUG-1999;  99US-0150566.
PR   25-AUG-1999;  99US-0150884.
PR   26-AUG-1999;  99US-0150804.
PR   27-AUG-1999;  99US-0151065.
PR   27-AUG-1999;  99US-0151066.
PR   27-AUG-1999;  99US-0151080.
PR   30-AUG-1999;  99US-0151303.
PR   31-AUG-1999;  99US-0151438.
PR   01-SEP-1999;  99US-0151930.
PR   07-SEP-1999;  99US-0152363.
PR   10-SEP-1999;  99US-0153070.
PR   13-SEP-1999;  99US-0153758.
PR   15-SEP-1999;  99US-0154018.
PR   16-SEP-1999;  99US-0154039.
PR   20-SEP-1999;  99US-0154779.
PR   22-SEP-1999;  99US-0155139.
PR   23-SEP-1999;  99US-0155486.
PR   24-SEP-1999;  99US-0155659.
PR   28-SEP-1999;  99US-0156458.
PR   29-SEP-1999;  99US-0156596.
PR   04-OCT-1999;  99US-0157753.
PR   05-OCT-1999;  99US-0157117.
PR   06-OCT-1999;  99US-0157865.
PR   07-OCT-1999;  99US-0158029.
PR   08-OCT-1999;  99US-0158232.
PR   12-OCT-1999;  99US-0158369.
PR   13-OCT-1999;  99US-0159293.
PR   13-OCT-1999;  99US-0159294.
PR   13-OCT-1999;  99US-0159295.
PR   14-OCT-1999;  99US-0159329.
PR   14-OCT-1999;  99US-0159330.
PR   14-OCT-1999;  99US-0159331.
PR   14-OCT-1999;  99US-0159637.
PR   14-OCT-1999;  99US-0159638.
PR   18-OCT-1999;  99US-0159584.
PR   21-OCT-1999;  99US-0160741.
PR   21-OCT-1999;  99US-0160767.
PR   21-OCT-1999;  99US-0160768.
PR   21-OCT-1999;  99US-0160770.
PR   21-OCT-1999;  99US-0160814.
PR   21-OCT-1999;  99US-0160815.
PR   22-OCT-1999;  99US-0160980.
PR   22-OCT-1999;  99US-0160981.
PR   22-OCT-1999;  99US-0160989.
PR   25-OCT-1999;  99US-0161404.
PR   25-OCT-1999;  99US-0161405.
PR   25-OCT-1999;  99US-0161406.
PR   26-OCT-1999;  99US-0161359.
PR   26-OCT-1999;  99US-0161360.
PR   26-OCT-1999;  99US-0161361.
PR   28-OCT-1999;  99US-0161920.
PR   28-OCT-1999;  99US-0161992.
PR   28-OCT-1999;  99US-0161993.
PR   29-OCT-1999;  99US-0162142.
PR   29-OCT-1999;  99US-0162144.
```

```
RESULT 27
AAU99928
ID   AAU99928 standard; Protein; 322 AA.
XX
AC   AAU99928;
XX
DT   07-OCT-2002  (first entry)
XX
DE   Human 84223 cation efflux predicted consensus sequence.
XX
KW   84223; cation efflux domain; infection; cancer;
KW   haematopoeitic disorder; blood clotting disorder;
KW   autoimmune disorder; leukaemia; immunological disorder;
KW   cardiovascular disorder; neurological disorder; cellular proliferation;
KW   red blood cell disorder; viral disease; neurological disorder.
OS   Synthetic.
XX
PN   WO200240656-A2.
XX
PD   23-MAY-2002.
XX
PF   14-NOV-2001; 2001WO-US45291.
XX
PR   14-NOV-2000; 2000US-248331P.
PR   14-NOV-2000; 2000US-248362P.
PR   14-NOV-2000; 2000US-248365P.
PR   30-NOV-2000; 2000US-250077P.
PR   30-NOV-2000; 2000US-250176P.
PR   30-NOV-2000; 2000US-250327P.
XX
PA   (MILL-) MILLENNIUM PHARM INC.
XX
PI   Meyers RE, Curtis RAJ, Glucksmann MA;
XX
DR   WPI; 2002-508325/54.
XX
PS   Disclosure; Fig 20; 298pp; English.
XX
CC   This invention relates to the DNA and protein sequences of novel
CC   isolated 47476, 67210, 49875, 46842, 33201, 83378, 84233, 64708, 85041
CC   or 84234 proteins. The method of the invention is useful for treating a
PT   Isolated 47476, 67210, 49875, 46842, 33201, 83378, 84233, 64708, 85041
PT   or 84234 polypeptides, useful as reagents or targets for treating or
PT   diagnosing pain or metabolic, liver, kidney, or cardiovascular
PT   disorders -
```

```
Qy    76  VASAICLLFMIGEVVGGYLAHSLAVWTDAAHLLTDFASMLISLFSLWMSSRPATKTMNFG  135
            : | :||||: | :|||||||||||||||| | :||| : ||: ||  |
Db    44  SLIFIYLIVMSVQIVGGFKANSLAVWTDAAHLLSDVAGLCVSLLAIKVSSWEANPRNSFG  103

Qy   136  WQRAELIGALLVSVLSIWVVTGVLLVYLLAVERLISGDYEIDGGTMLITSGCAVANIIMGLT  195
            :| | :| : |:|:| :|:| :|||||:|||::| : ||| ||||:||| |:| :|::|
Db   104  FKRLEVLAAFLSVQLIWLVSGVIIHEAIQRLLSRSREVNGEIMFGISAFGFFMNLVMVLW  163

Qy   196  LHQSGHGHSH----------------------GTTNQQEENP-----------SVR  218
            | :|| ||                          ||                  : :
Db   164  L---GHNHSHHHDHHHHHHHHHHHKHQHHHHKEVVAEEEEEMNPLKGEKSSSKEMNINIQ  220

Qy   219  AAFIHVIGDFMQSMGVLVAAYILYFKPEYKYVDPICTFVFSILVLGTTLTIILRDVILVLM  278
            | :|  : ||:::||  | |: |||| | ||  ||  || |||| |  ||: | |||:||
Db   221  GAYLHAMADMIQSLGVMIGGGIIWVKPKWVLVDLICTIVFSAFALAAATLPILKNIFGILM  280

Qy   279  EGTPKGVDFTAVRDLLLSVEGVEALHSLHITWALTVAQPVLSVHIAIAQNTDAQVLKTAS  338
            | |  | |:| |||  :| | |:|| |||  || |     :|  :  : |   |  |
Db   281  ERVPRDMDIEKLERGLKRIDGVKIVYDLHWEITVGRIVLSCHLPEPGASPKEIITGVR  340

Qy   339  SRLQGKFHFHTVTIQIE  355
            :  | |  ||  |:| |
Db   341  NFCRKSYGIYHATVQVE  357
```

```
Query Match  30.0%;  Score 498.5;  DB 21;  Length 359;
Best Local Similarity  25.9%;                       Pred. No. 8.5e-45;
Matches 113;  Conservative 81;  Mismatches 126;  Indels 57;  Gaps 5;

Qy    16  IRSYTGSLWQEGAGWIPLPRPGLDLQAIELAAQSNHHCHAQKGPDSHCDPKKGKAQRQLY  75
            : | :| : | : || |  | ||              : :|| :  : :||  | :|| :|
Db     1  MESPSPSKTEENLGVVPL------------------SCAFTRQEHCVSE------------TKEREESTRRLS  43
```

CC disorder characterised by aberrant activity of 47476, 67210, 49875,
CC 46842, 33201, 83378, 84233, 64708, 85041 or 84234-expressing cell in a
CC subject. The protein molecules can act as novel diagnostic targets and
CC therapeutic agents for controlling aberrant or deficient signal
CC transduction resulting, in e.g., haematopoeitic disorders, including
CC blood clotting disorders, autoimmune disorders, or disorders related to
CC an inability to clear infections (e.g., viral or bacterial infections),
CC as well as disorders related to abnormal cellular proliferation or
CC differentiation, e.g., leukaemia. They may also be used to control
CC disorders of metabolic imbalance (e.g., disorders of lipopolysaccharide
CC biosynthesis or glycogen synthesis) immunological disorders,
CC cardiovascular disorders, neurological disorders, or cellular
CC proliferation and/or differentiation disorders, e.g., cancer, cell
CC motility and adhesion disorders differentiative disorders, e.g., brain
CC cell disorders, viral diseases, neurological disorders (e.g., brain
CC disorders), pain or metabolic disorders, liver disorders, kidney
CC disorders, disorders of the small intestine, disorders of metal ion
CC imbalance, protein trafficking disorders and disorders associated with
CC bone metabolism. The sequences of the invention are also useful for
CC screening assays, predictive medicine (e.g., diagnostic assays,
CC prognostic assays, monitoring clinical trials, and pharmacogenetics);
CC and methods of treatment (e.g., therapeutic and prophylactic). The
CC present sequence represents a predicted consensus sequence motif
CC found in the protein of the invention.
XX
SQ Sequence 322 AA;

Query Match   24.4%;   Score 469;   DB 23;   Length 322;
Best Local Similarity   34.5%;   Pred. No. 1.1e-41;
Matches   110;   Conservative   70;   Mismatches   105;   Indels   34;   Gaps   6;

Qy 76 VASAICLLFMIGEVVGGYLAHSLAVMTDAAHLLTDFASMLISLFSLWMSSRPATKTMNFG 135
Db 3 ISLAINLLMLIKLIGGVLSGSLALLADALHSLSDVASSLISLIALRLAEKPDDEKHPFG 62

Qy 136 WQRAEILGALVSVLSIWVVTGV-LVYLAVERLISGDYEID----------------- 174
Db 63 HHRAETLAALLNSVFLVIVSFLEILYEAIERLISPDYEIPPDAVLAADIMEPEEPGLFEV 122

Qy 175 GGTMLITSGCAVAVNIMGLTLHQSGHGHSH--GTTNQQEENPSVRAAFIHVIGDFMQSM 232
Db 123 GGVALGVALGGTALVVLLGLVVNLALHGYLRRVGKKLKSEHNLNVRAAALHVLGDALSSV 182

Qy 233 GVLVAAYILYFK-------PEYKYDPICIFVFSILVLGTTLTLIRDVILVLMEGTPKGV 285
Db 183 GVLIRAALLIYFTGYSFKGWKKWYYADPIASILISLLILYTAFRLKESVLLLEGTPSKE 242

Qy 286 DF-TAVRDLLLSVEGVEALHSLHIWALTVAQPVLSVHIAIAQNTD---AQAVLKTASSRL 341
Db 243 DLERKIKKTLLSIPGVKGVHDLHIWLGSNKFIASVHVEVDDNLDLKEAHDILAEIEREI 302

Qy 342 QGKFHFHTVTIQIEDYSED 360
Db 303 LHKFGIEHVTHVEPASEE 321

RESULT 28
AAU99931
ID   AAU99931   standard;   Protein;   322 AA.
XX
AC   AAU99931;
XX
DT   07-OCT-2002   (first entry)
XX
DE   Human 85041 cation efflux predicted consensus sequence.
XX
KW   85041; cation efflux domain; infection; cancer;
KW   haematopoetic disorder; blood clotting disorder;
KW   autoimmune disorder; leukaemia; immunological disorder;
KW   cardiovascular disorder; neurological disorder; cellular proliferation;
KW   red blood cell disorder; viral disease; neurological disorder.
XX
OS   Synthetic.
XX
PN   WO200240656-A2.
XX
PD   23-MAY-2002.
XX
PF   14-NOV-2001; 2001WO-US45291.
XX
PR   14-NOV-2000; 2000US-248331P.
PR   14-NOV-2000; 2000US-248362P.
PR   14-NOV-2000; 2000US-248365P.
PR   30-NOV-2000; 2000US-250176P.
PR   30-NOV-2000; 2000US-250177P.
PR   30-NOV-2000; 2000US-250327P.
XX
PA   (MILL-) MILLENNIUM PHARM INC.
XX
PI   Meyers RE, Curtis RAJ, Glucksmann MA;
XX
DR   WPI; 2002-508325/54.
XX
PT   Isolated 47476, 67210, 49875, 46842, 33201, 83378, 84233, 64708, 85041
PT   or 84234 polypeptides, useful as reagents or targets for treating or
PT   diagnosing pain or metabolic, liver, kidney, or cardiovascular
PT   disorders.
XX
PS   Disclosure; Fig 26; 298pp; English.
XX

RESULT 29

```
Qy   233 GVLVAAYILYFK------------PEYKYVDPICTFVFSILVLGTTLTIILRDVILVLMEGTPKGV 285
             |||:||  ::|||               |:|||  |:::|||:|:::||:|||:::|||:|||||
Db   183 GVLIAALLIYFTGYSFKGWKWWYYADPIASILISLIILYTAFRLLKESVLILLEGTPSKE 242

Qy   286 DF-TAVRDLLLSVEGVEALHSLHIWALTVAQPVLSVHIAIAQNTD---AQAVLKTASSRL 341
             ::  ||||||||||||:|||:|| :|  ||||| |  ||:|| |       |:|||:|
Db   243 DLERKIKTLLSIPGVKGVHDLHIWYLGSNKFIASVHVEVDDNLDLKEAHDILAEIEREI 302
```

Qy   342 QGKFHFHTVTIQIEDYSED 360
             || ||:| ::| ||:
Db   303 LHKFGIEHVTVHVEPASEE 321

```
CC       motility and adhesion disorders differentiative disorders, red blood
CC       cell disorders; viral diseases, neurological disorders (e.g., brain
CC       disorders), pain or metabolic disorders, liver disorders, kidney
CC       disorders, disorders of the small intestine, disorders of metal ion
CC       imbalance, protein trafficking disorders and disorders associated with
CC       bone metabolism. The sequences of the invention are also useful for
CC       screening assays, predictive medicine (e.g., diagnostic assays,
CC       prognostic assays, monitoring clinical trials, and pharmacogenetics);
CC       and methods of treatment (e.g., therapeutic and prophylactic). The
CC       present sequence represents a predicted consensus sequence motif
CC       found in the protein of the invention.
XX
SQ       Sequence   289 AA;
```

RESULT 30

```
Query Match             23.0%;  Score 442.5;  DB 23;  Length 289;
Best Local Similarity   35.4%;  Pred. No. 6.7e-39;
Matches 101;  Conservative  65;  Mismatches 88;  Indels 31;  Gaps 5;

Qy    76 VASAICLLFMIGEVVGGYLAHSLAVMTDAAHLLTDFASMLISLFSIWMSSRPATKTMNFG 135
             :: |:||::| |:::||| |:::|::| |::||:|:|:| |:||::| :|:|| |  ||
Db     3 ISLALNILLMLIKLIGGVLSGSIALLADALHSLSDVASSLISLLALRLAEKPPDEKHPFG 62

Qy   136 WQRAEILGALVSVLSIWVVTGV-LVYLAVERLISGDYEID----------------------- 174
          ||| || |:| || | : ::  |: :|| ||||| |:||
Db    63 HHRAETLAALLNSVFLVIVSFLEILYEAIERLISPDYEIPPDAVLAADIMEPEEPGLFEV 122

Qy   175 GGTMLITSGCAVANVNIIMGLTLHQSGHGHSH--GTTNQOEENPSVRAAFIHVIGDFMQSM 232
          |:  ::  | |::||  | | ||:|:|:    :|:||| |:|:|| |:|:|| :  |:
Db   123 GGVALGVALGGTALVVLLGLVNNLALHGTLRVGKKLKSEHNLNVRAAALHVLGDALSSV 182

Qy   233 GVLVAAYILYFK------------PEYKYVDPICTFVFSILVLGTTLTIILRDVILVLMEGTPKGV 285
             ||||:|  ::|||               |:|||  |:::|||:|:::||:|||:::||||||
Db   183 GVLIAALLIYFTGYSFKGWKWWYYADPIASILISLIILYTAFRLLKESVLILLEGTPSKE 242

Qy   286 DF-TAVRDLLLSVEGVEALHSLHIWALTVAQPVLSVHIAIAQNTD 329
             ::  ||||||||||||:|||:|| :|  ||||| |  ||:|| |
Db   243 DLERKIKTLLSIPGVKGVHDLHIWYLGSNKFIASVHVEVDDNLD 287
```

RESULT 30

```
ID   ABB53142 standard; Protein; 144 AA.
XX
AC   ABB53142;
XX
DT   12-FEB-2002  (first entry)
XX
DE   Human ORF48 protein.
XX
OS   Homo sapiens.
XX
XX
XX   WO200177155-A2.
XX
PN   WO200177155-A2.
XX
PD   18-OCT-2001.
XX
PF   06-APR-2001; 2001WO-US040469.
PR   06-APR-2000; 2000US-195576P.
PR   05-APR-2001; 2001US-0826734.
XX
PA   (CURA-) CURAGEN CORP.
XX
PI   Fernandes E, Vernet CAM, Mishnu VS, Leach MD, Shimkets RA;
```

```
DE   Human 84234 cation efflux predicted consensus sequence.
XX
KW   84234; cation efflux domain; infection; cancer;
KW   haematopoeitic disorder; blood clotting disorder;
KW   autoimmune disorder; leukaemia; immunological disorder;
KW   cardiovascular disorder; neurological disorder; cellular proliferation;
KW   red blood cell disorder; viral disease; neurological disorder.
OS   Synthetic.
XX
PN   WO200240656-A2.
XX
PD   23-MAY-2002.
XX
PF   14-NOV-2001; 2001WO-US45291.
XX
PR   14-NOV-2000; 2000US-248331P.
PR   14-NOV-2000; 2000US-248362P.
PR   14-NOV-2000; 2000US-248365P.
PR   30-NOV-2000; 2000US-250077P.
PR   30-NOV-2000; 2000US-250176P.
PR   30-NOV-2000; 2000US-250327P.
XX
PA   (MILL-) MILLENNIUM PHARM INC.
XX
PI   Meyers RE, Curtis RAJ, Glucksmann MA;
XX
DR   WPI; 2002-508325/54.
XX
PT   Isolated 47476, 67210, 49875, 46842, 33201, 83378, 84233, 64708, 85041
PT   or 84234 polypeptides, useful as reagents or targets for treating or
PT   diagnosing pain or metabolic, liver, kidney, or cardiovascular
PT   disorders
XX
PS   Disclosure; Fig 29; 298pp; English.
XX
CC   This invention relates to the DNA and protein sequences of novel
CC   isolated 47476, 67210, 49875, 46842, 33201, 83378, 84233, 64708, 85041
CC   or 84234 proteins. The method of the invention is useful for treating a
CC   disorder characterised by aberrant activity of 47476, 67210, 49875,
CC   46842, 33201, 83378, 84233, 64708, 85041 or 84234-expressing cell in a
CC   subject. The protein molecules can act as novel diagnostic targets and
CC   therapeutic agents for controlling aberrant or deficient signal
CC   transduction resulting, in e.g., haematopoeitic disorders, including
CC   blood clotting disorders, autoimmune disorders, or disorders related to
CC   an inability to clear infections (e.g., viral or bacterial infections),
CC   as well as disorders related to abnormal cellular proliferation or
CC   differentiation, e.g., leukaemia. They may also be used to control
CC   disorders of metabolic imbalance (e.g., disorders of lipopolysaccharide
CC   biosynthesis or glycogen synthesis) immunological disorders,
CC   cardiovascular disorders, neurological disorders,
CC   disorders of metabolic imbalance (e.g., disorders of lipopolysaccharide
CC   biosynthesis or glycogen synthesis) immunological disorders,
CC   proliferation and/or differentiation disorders, e.g., cancer, cell
```

```
ORFX: human; pharmacogenomics; cancer; hyperproliferative disorder;
dysproliferative disorder; neurodegenerative disorder; organ transplant;
cardiovascular disease; cytokine; cell proliferation; immunomodulatory;
cell differentiation; haematopoiesis; tissue growth; thrombolytic;
nootropic; neuroprotective; antiarthritic; antimicrobial; vulnerary;
cytostatic; antidiabetic; virucide; antiinfertility; anticonvulsant;
vasotropic; antiparkinsonian; dermatological; antiinflammatory;
antiulcer; osteopathic; tranquilliser; cerebroprotective; hepatotropic;
antiinflammatory; tumour inhibition.
```

PI Zerhusen BD, Kekuda R;
XX
DR WPI; 2002-049131/06.
DR N-PSDB; ABA90241.
XX
PS Claim 11; Page 90; 111pp; English.

PT New polypeptides and polynucleotides, useful for treating diseases such
PT as cancer, Alzheimer's disease, atherosclerosis, diseases associated
PT with liver, comprises polypeptides and polynucleotides of open reading
PT frame proteins -
XX

CC The present invention provides the protein and coding sequences of human
CC ORFX, where X is any number between 1 and 132. These sequences can be
CC used to diagnose and treat ORFX related disorders, including cancer,
CC hyperproliferative and dysproliferative disorders, neurodegenerative
CC disorders, disorders related to organ transplantation or cardiovascular
CC diseases, and may have cell proliferation or differentiation, cytokine,
CC immunomodulatory, haematopoiesis regulating, tissue growth, inhibin or
CC activin, chemotactic or chemokinetic, haemostatic or thrombolytic
CC and/or antiinflammatory activities. The present sequence is one of the
CC ORFX proteins of the invention.
XX
SQ Sequence 144 AA;

Query Match 59.2%; Score 430; DB 23; 2e-38;
Best Local Similarity 59.2%; Pred. No. 5,
Matches 84; Conservative 21; Mismatches 37; Indels 0; Gaps 0;

```
Qy  51 HHCHAQKGPDSHCDPKKGKRAQRQLYVASAICLLFMIGEVGVGGYLAHSLAVWTDAAHLLTD 110
        ::||:| |||||:| |:|::|||||||:|:|||||||||||||||:||||||||||||||
Db   1 YHCHSGSKPTEKGANEYAYAXWKLCSASAICFIFMIAEVVGGHIAGSLAVVTDAAHLLID  60

Qy 111 FASMLISLFSLWMSSRPATKTMNFGWQRAEILGALVSVLSIWVVTGVLVIVLAVERLISGD 170
        |:|||||||||||||||||:||||||||||||||||||||||:|||||||:||:  |
Db  61 LTSFLLSLFSLWLSSKPPSKRLTFGWHRAEILGALLSILCIWVVTGVLVYLACERLLYPD 120

Qy 171 YEIDGGTMLITSGCAVAVNIIM 192
        |:| |||| |||| ||:|| |:
Db 121 YQIQATVMIIVSSCAVAANIVL 142
```

RESULT 31
ABB48454
ID ABB48454 standard; Protein; 303 AA.
XX
AC ABB48454;
XX
DT 05-FEB-2002 (first entry)
XX
DE Listeria monocytogenes protein #1158.
KW Antibacterial; gene therapy; vaccine; biosynthesis; biodegradation;
KW vitamin B12; bacterial infection; disease.
OS Listeria monocytogenes.
XX
PN WO200177335-A2.
XX
PD 18-OCT-2001.
XX
PF 11-APR-2001; 2001WO-FR01118.
XX
PR 11-APR-2000; 2000FR-0004629.
XX
PA (INSP) INST PASTEUR.
XX
PI Buchrieser C, Frangeul L, Couve E, Rusniok C, Fsihi H, Dehoux P;
PI Dussurget O, Chetouani F, Nedjari H, Glaser P, Kunst F, Cossart P;
PI Daniels J, Goebel W, Kreft J, Kuhn M, Ng E, Vazquez-Boland JA;
PI Dominguez-Bernal G, Garrido-Garcia P, Tierrez-Martinez A, Amend A;
PI Chakraborty T, Domann E, Hain T, Berche P, Charbit A, Durant L;
PI Perez-Diaz J, Baquero F, Garcia Del Portillo F, Gomez-Lopez N;

PI Maduenio E, De Pablos B, Wehland J, Kaerst U, Entian K, Hauf J;
PI Rose M, Voss H;
XX
DR WPI; 2002-010914/01.
XX
PT Genomic sequence for Listeria monocytogenes, useful e.g. for treatment
PT and prevention of Listeria and related bacterial infections, and
PT related polypeptides -
XX
PS Claim 6; SEQ ID No 1159; 192pp; French.

CC The present invention relates to the genome sequence of Listeria
CC monocytogenes EGD-e (see AAA03041). The genome sequence and fragments of
CC it are useful for selecting probes and primers for detecting genes in L.
CC monocytogenes and related organisms, and for studying genetic
CC polymorphisms and other genomes. The present sequence is a protein
CC encoded by the genome sequence of the present invention. Proteins
CC expressed from the genome sequence are useful for raising specific
CC antibodies. Identification of L. monocytogenes and related organisms, and
CC for biosynthesis and biodegradation, especially biosynthesis of Vitamin
CC B12. The genome sequence and proteins encoded by it are also useful for
CC selecting compounds that regulate gene expression and cell replication
CC and modulate L. monocytogenes-related diseases. In addition, the genome
CC sequence and proteins encoded by it are useful in pharmaceutical and
CC vaccines compositions for the treatment or prevention of infections by L.
CC monocytogenes and related organisms.
CC Note: The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ Sequence 303 AA;

Query Match 30.1%; Score 414.5; DB 23; 7.4e-36;
Best Local Similarity 30.1%; Pred. No. 7,
Matches 93; Conservative 72; Mismatches 127; Indels 17; Gaps 5;

```
Qy  47 AQSNHHCHAQKGPDSHCDPKKGKAQRQLYVASAICLLFMIGEVGVGGYLAHSLAVWTDAAH 106
        | | ||   |            :||||||||||:|||||||:|:||||||||||||||
Db   2 AHNHDAHGHNHNHAHNANKK------SLFISFLLIATFMVVEVIGGIMTNSLALLSDAGH  56

Qy 107 LLTDFASMLISLFSLWMSSRPATKTMNFGWQRAEILGALVSVLSIWVVTGVLVYLAVERL 166
        :|:| | :|| |:| |::    :||  : : |||:| ||: |:::||||||| |||||
Db  57 MLSDAVALGLSLAAFKFGEKAASSDKTYGYKRFEILAAFLNGLTLVGISVFFYEAIGRF 116

Qy 167 ISGDYEIDGGTMLITSGCAVAVNIIMGLTLHQSGHGHSHGTTNQQEENPSYRRAAFIHVIG 226
        :  |  :  | ::||  : :::| :|||::      |: :|      | : |:||||:|:|
Db 117 FDPPQVIGAGMMTI--SVIGLLINILVAWILMK------GD?S---ENLNMRSAFLHVLG 165

Qy 227 DPMQSMGVLVAAYILYFKPEYKYVDPICTFVFSILVLGTTLTLIRDVILVLMEGTPKGVD 286
        | : | | ::||  | |     :  || :|::|:||:||||| | :| | |||| |  ||
Db 166 DLLGSVGAIIAALLIFL-GWNIADPIASVIVAALILVSGWRVLKDAIHILMEGKPANVD 224

Qy 287 FTAVRDLLLSVEGVEALHSLHIWALTVAQPVLSVHIAIAQNIDAQVLKTASSRLQGKFH 346
        |  :::||: :   |:|| ||:| |:|    |:: |    | | : :     :|| ::
Db 225 TEEIKTFFQQQDGVKEVHDLHVWAITSDFNALSAHLTVCEDADRDKILADIEHYLQENFS 284

Qy 347 FHTVTIQIE 355
        :  |||:|
Db 285 LEHSTIQIE 293
```

RESULT 32
ABP39345
ID ABP39345 standard; Protein; 342 AA.
XX
AC ABP39345;
XX
DT 24-JUL-2002 (first entry)
XX
DE Staphylococcus epidermidis ORF amino acid sequence SEQ ID NO:4190.
KW Staphylococcus epidermidis; open reading frame; ORF; bacterial infection;
KW antibacterial; gene therapy.

RESULT 33
ABP40773
ID   ABP40773 standard; Protein; 359 AA.
XX
AC   ABP40773;
XX
DT   24-JUL-2002 (first entry)
XX
DE   Staphylococcus epidermidis ORF amino acid sequence SEQ ID NO:5618.
XX
KW   Staphylococcus epidermidis; open reading frame; ORF; bacterial infection;
KW   antibacterial; gene therapy.
XX
OS   Staphylococcus epidermidis.
XX
PN   US6380370-B1.
XX
PD   30-APR-2002.
XX
PF   13-AUG-1998;   98US-0134001.
XX
PR   14-AUG-1997;   97US-055779P.
PR   08-NOV-1997;   97US-064964P.
XX
PA   (GENO-) GENOME THERAPEUTICS CORP.
XX
PI   Doucette-Stamm LA, Bush D;
XX
DR   WPI; 2002-381255/41.
DR   N-PSDB; ABN93318.
XX
PT   Novel isolated nucleic acid encoding a Staphylococcus epidermidis
PT   polypeptide, useful for diagnosing and treating bacterial infections -
XX
PS   Disclosure; SEQ ID 5618; 267pp; English.
XX
CC   ABN90538 to ABN93374 represent Staphylococcus epidermidis open reading
CC   frame (ORF) nucleic acid sequences which encode the amino acid sequences
CC   given in ABP35124 to ABP37960. The S. epidermidis sequences have
CC   antibacterial activity and can be used in gene therapy. The sequences
CC   can also be used in the diagnosis and treatment of bacterial infections,
CC   particularly S. epidermidis infections. The sequences can be used to
CC   screen for compounds able to interfere with the S. epidermidis life
CC   cycle or inhibit S. epidermidis infection.
CC   N.B. The sequence data for this patent did not form part of the printed
CC   specification, but was obtained in electronic format directly from the
CC   USPTO web site.
XX
SQ   Sequence   359 AA;

Query Match   20.0%; Score 385; DB 23; Length 359;
Best Local Similarity   27.7%; Pred. No. 1.5e-32;
Matches   88; Conservative   89; Mismatches 119; Indels   22; Gaps   7;

Qy   50 NHHCHA------QKGPD--SHCDPKKGKAQRQ--LYVASAICLLFMIGEVVGGYLAHSLAV 100
                | | | :| | | :   :||||||| ||| ||| ||| ||
Db   37 NRRIHAWRRDMENKPDYFHIHEHRKFQSSSKITLWLSLVITIMIFVVEFVGGIVSNSLAL 96

Qy   101 MTDAAHLLTDFASMLISLFSLWMSSRPATKTMNFGWQRAEILGALVSVLSIWVVTGVYLVLY 160
               :|| |||||||||||||||||||||||||||||:|||||||||| |||| |: || | |
Db   97 LSDSFHMLSDVLALGLSMVAIYFSSKPPTKNYTGFLRLEIIVAFLNGLALIVISLGIMY 156

Qy   161 LAVERLISGDYEIDGGTMLITSGCAVAVNIIMGLTLHQSGHGHSGTNQQEENPSVRAA 220
               |: | ||| | ||| :|| |||| ||| | || || | | :: |
Db   157 EGIMRIIH-PRPVESGIMILIAFIGLIANIVLTIIIMIS------LKKENNINIQSA 206

Qy   221 FIHVIGDFMQSMGVLVAAYILYFKPEYKYVDPICTFVFSILVLGTTLTILRDVILVLMEG 280
                | | ||:|||| :| | | | |:|| | :| :| :| || | |||: | |
Db   207 LWHFIGDLLNSLGIIVAFVLIHF-TGWNIVDPIISILISLIILRGGYKIIKNASKVLMER 265

Qy   281 TPKGVDFTAVRDLLLSVEGVEALHSLHIWALTVAQPVLSVHIAIAQN-TDAQAVLKTAS 338
                | |||| | || |:| ||| | :|: || | ::: :||| ||| |:| |
Db   266 VPDRYDTDEIMGAMKDVEGVIDIHEFHLMSVTTNQSSLSAHVVLSDDYIKSPYAINKVS 325

Qy   339 SRLQGKFHFHTVTIQIED 356
                |:|:| |||| |:|| |
Db   326 DLLKTQYGLEHVTLQIEN 343

XX
OS   Staphylococcus epidermidis.
XX
PN   US6380370-B1.
XX
PD   30-APR-2002.
XX
PF   13-AUG-1998;   98US-0134001.
XX
PR   14-AUG-1997;   97US-055779P.
PR   08-NOV-1997;   97US-064964P.
XX
PA   (GENO-) GENOME THERAPEUTICS CORP.
XX
PI   Doucette-Stamm LA, Bush D;
XX
DR   WPI; 2002-381255/41.
DR   N-PSDB; ABN91890.
XX
PT   Novel isolated nucleic acid encoding a Staphylococcus epidermidis
PT   polypeptide, useful for diagnosing and treating bacterial infections -
XX
PS   Disclosure; SEQ ID 4190; 267pp; English.
XX
CC   ABN90538 to ABN93374 represent Staphylococcus epidermidis open reading
CC   frame (ORF) nucleic acid sequences which encode the amino acid sequences
CC   given in ABP35124 to ABP37960. The S. epidermidis sequences have
CC   antibacterial activity and can be used in gene therapy. The sequences
CC   can also be used in the diagnosis and treatment of bacterial infections,
CC   particularly S. epidermidis infections. The sequences can be used to
CC   screen for compounds able to interfere with the S. epidermidis life
CC   cycle or inhibit S. epidermidis infection.
CC   N.B. The sequence data for this patent did not form part of the printed
CC   specification, but was obtained in electronic format directly from the
CC   USPTO web site.
XX
SQ   Sequence   342 AA;

Query Match   21.2%; Score 408.5; DB 23; Length 342;
Best Local Similarity   28.4%; Pred. No. 4e-35;
Matches   97; Conservative   80; Mismatches 144; Indels   21; Gaps   7;

Qy   32 PLPRPGDLDLQAIELAAQSNH-HCHAQRGGPDSHCDPKKGKGKAQRQLYVASAICLLFMIGEVV 90
                 | :: :| | | | :|| | : | :   :||||| ||
Db   12 PFTTPNILVKVEFLMTNHNHQHHHSHAHGHVHTDNKK------VLMFSFIIISLFMLVEII 66

Qy   91 GGYLAHSLAVWTDAAHLLTDFASMLISLFSLWMSSRPATKTMNFGWQRAEILGALVSVLS 150
                |:||||||||| ||||||:||| :| || || : |||||||||||||| |||| |
Db   67 GGFLANSLALLSDGFHMLSDAISLGVALIAFIYAEKHATKSKTYGYKRFEILAALFNGVT 126

Qy   151 IWVVTGVYLVLYLAVERLISGDYEIDGGTMLITSGCAVAVNIIMGLTLHQSGHGHSGTNQ 210
                |:|: ||| : || | |:| | || :| |||| ||| | ||| || | |
Db   127 LFIISIIIIIEAIRRFLEPP-EVQSKEMFIISVIGLMVNIIVAILMFKGG------D 176

Qy   211 QEENPSVRAAFIHVIGDFMQSMGVLVAAYILYFKPEYKYVDPICTFVFSILVLGTTLTIL 270
                 :: : | |||| :| | : :|:| | | | | | :| | :: :| :| || |
Db   177 TSHNLNRGAFLHVLGDLFGSVGAIVAS-LLIWGFNFIADPIASILVSLIILKSAYGIS 235

Qy   271 RDVILVLMEGTFKGVDFTAVRDLLLSVEGVEALHSLHIWALTVAQPVLSVHIAIAQNTDA 330
                 : ||:|||| :| || | | |:| || | : :|| || | : ::||||
Db   236 KSSLNILMEGTPNDIDLNAVIRAISKDERIQNVHDCHVWTISNDMNALSCHAVVPEYLSV 295

Qy   331 Q--AVLKTASSRLQGKFHFHTVTIQIEDYSEDMKDCQOACQG 369
                :   :: ||| ::| || |
Db   296 QTCETMLKSIESDLL-QLNIQHMTIQLETPEHKHDESTLCSG 336

RESULT 34

ABB57822

ID   ABB57822 standard; Protein; 449 AA.
XX
AC   ABB57822;
XX
DT   26-MAR-2002 (first entry)
XX
DE   Drosophila melanogaster polypeptide SEQ ID NO 258.
XX
KW   Drosophila; developmental biology; cell signalling; insecticide;
KW   pharmaceutical.
XX
OS   Drosophila melanogaster.
XX
PN   WO200171042-A2.
XX
PD   27-SEP-2001.
XX
PF   23-MAR-2001; 2001WO-US09231.
XX
PR   23-MAR-2000; 2000US-1916337P.
PR   11-JUL-2000; 2000US-0614150.
XX
PA   (PEKE) PE CORP NY.
XX
PI   Venter JC, Adams M, Li PWD, Myers EW;
XX
PT   New isolated nucleic acid detection reagent for detecting 1000 or more
PT   genes from Drosophila and for elucidating cell signalling and cell-cell
PT   interactions -
XX
PS   Disclosure; SEQ ID NO 258; 21pp + Sequence Listing; English.
XX
CC   The invention relates to an isolated nucleic acid detection reagent
CC   capable of detecting 1000 or more genes from Drosophila. The invention is
CC   useful in developmental biology and in elucidating cell signalling and
CC   cell-cell interactions in higher eukaryotes for the development of
CC   insecticides, therapeutics and pharmaceutical drugs. The invention
CC   discloses genomic DNA sequences (ABL01840-ABL16175) and the encoded proteins
CC   sequences (ABL01840-ABL16175) and the encoded proteins
CC   (ABB57737-ABB72072).
CC   The sequence data for this patent did not form part of the printed
CC   specification, but was obtained in electronic format directly from WIPO
CC   at ftp.wipo.int/pub/published_pct_sequences.
XX
DR   WPI; 2001-656860/75.
DR   N-PSDB; ABL01925.
XX
SQ   Sequence   449 AA;

Query Match   19.0%; Score 366; DB 22; Length 449;
Best Local Similarity   28.9%; Pred. No. 2.3e-30;
Matches 103; Conservative 74; Mismatches 117; Indels 62; Gaps 14;

Qy   68 GKAQRQLYVASAICLLFMIGEVVGGYLAHSLAVMTDAAHLLTDFASMLISLFSLWMSSRP 127
          :| ||| |:|:::| || :|| | |:|:| |||::||| | :::||:| || ::|| |
Db   6 GKKCRLLSMWWLTAFFFV-EIIVGYVTNSMALVADSFHMLGDIAALVISFLSVKMSPKK 64

Qy   128 ATKTMNFGWQRAEILGALVSVLSIWVVTGVLVYLYLAVERLISGDYEIDGGTMLITSGCAVA 187
          ||||| ||| | | |||| || || | ||| :| | || :| | | ||| :|
Db   65 WSKN-TFGWARAEVLGALVNAVFLVALCFSITIEACKRFIEEEPIHEPELLIVGALGLL 123

Qy   188 VNIIMGLTL---HQSGHGHSHG---TTN---------------------QQEENP-- 215
          ||||||| | ||||||||| |:::| :|:| |
Db   124 VNVI-GLCLLYEHGGHHGHSHGGGLTRNHSRLTELANMDEGEDEQNDFAYEKQKEKAPVK 182

Qy   216 -------SVRAAFIHVIGDFMQSMGVLVAAVILYFKPEYK---YVDPICTFVFS 259
          :| || |:| | |:| || | :| |: | :::| |||::|:
Db   183 KSSHGHSHDPGQMMRGAFLHVLSDALGSIIVVISAVAVVW-KTEWKYRYYMDPALSIVLV 241

Qy   260 ILVLGTTLTITILRDVILVLIMEGTPKGVDFTAVRDLLL-SVEGVEALHSLHIWALTVAQPVL 318
          :|:|: | | |||: ||| :||| | : ||: |
Db   242 VLILHSVWPLLRESALILQTVPTHIQVDAIQKRLLEKVDGVLAVHEPHVWQLAGDRIIA 301

Qy   319 SVHIAIAQNTDAQAVLKTASSRLQGKFH---FHVTIQIEDYSEBDMKDCQACQGPS 371
          :: | ||| |:| ||| | | || |||:| | |||:: :| ::| |
Db   302 SAHIRCRNLSEYMKI----AEKVKEFFHNEGIHSTTIQPE-FSE-IEGCNMSDGTS 351

RESULT 35

ABB66435

ID   ABB66435 standard; Protein; 449 AA.
XX
AC   ABB66435;
XX
DT   26-MAR-2002 (first entry)
XX
DE   Drosophila melanogaster polypeptide SEQ ID NO 26097.
XX
KW   Drosophila; developmental biology; cell signalling; insecticide;
KW   pharmaceutical.
XX
OS   Drosophila melanogaster.
XX
PN   WO200171042-A2.
XX
PD   27-SEP-2001.
XX
PF   23-MAR-2001; 2001WO-US09231.
XX
PR   23-MAR-2000; 2000US-1916337P.
PR   11-JUL-2000; 2000US-0614150.
XX
PA   (PEKE) PE CORP NY.
XX
PI   Venter JC, Adams M, Li PWD, Myers EW;
XX
PT   New isolated nucleic acid detection reagent for detecting 1000 or more
PT   genes from Drosophila and for elucidating cell signalling and cell-cell
PT   interactions -
XX
PS   Disclosure; SEQ ID NO 26097; 21pp + Sequence Listing; English.
XX
CC   The invention relates to an isolated nucleic acid detection reagent
CC   capable of detecting 1000 or more genes from Drosophila. The invention is
CC   useful in developmental biology and in elucidating cell signalling and
CC   cell-cell interactions in higher eukaryotes for the development of
CC   insecticides, therapeutics and pharmaceutical drugs. The invention
CC   discloses genomic DNA sequences (ABL01840-ABL16175) and the encoded proteins
CC   sequences (ABL01840-ABL16175) and the encoded proteins
CC   (ABB57737-ABB72072).
CC   The sequence data for this patent did not form part of the printed
CC   specification, but was obtained in electronic format directly from WIPO
CC   at ftp.wipo.int/pub/published_pct_sequences.
XX
DR   WPI; 2001-656860/75.
DR   N-PSDB; ABL10538.
XX
SQ   Sequence   449 AA;

Query Match   19.0%; Score 366; DB 22; Length 449;
Best Local Similarity   28.9%; Pred. No. 2.3e-30;
Matches 103; Conservative 74; Mismatches 117; Indels 62; Gaps 14;

Qy   68 GKAQRQLYVASAICLLFMIGEVVGGYLAHSLAVMTDAAHLLTDFASMLISLFSLWMSSRP 127
          :| ||| |:|:::| || :|| | |:|:| |||::||| | :::||:| || ::|| |
Db   6 GKKCRLLSMWWLTAFFFV-EIIVGYVTNSMALVADSFHMLGDIAALVISFLSVKMSPKK 64

Qy   128 ATKTMNFGWQRAEILGALVSVLSIWVVTGVLVYLYLAVERLISGDYEIDGGTMLITSGCAVA 187
          ||||| ||| | | |||| || || | ||| :| | || :| | | ||| :|
Db   65 WSKN-TFGWARAEVLGALVNAVFLVALCFSITIEACKRFIEEEPIHEPELLIVGALGLL 123

Qy   188 VNIIMGLTL---HQSGHGHSHG---TTN--------------------- 215
          ||||||| | ||||||||| |:::|
Db   124 VNVI-GLCLLYEHGGHHGHSHGGGLTRNHSRLTELANMDEGEDEQNDFAYEKQKEKAPVK 182

Left column

```
Qy   216 ----------SVRAAFIHVIGDFMQSMGVLVAAYILYFKPEYK---YVDICTFVFS 259
                   ::|||   ||||:  ::|||:||||:||||||||     ||:||  |:
Db   183 KSSHGHSHDPGQMMRGAFLHVLSDALGSSIIVVISAVVVW-KIEWKIRYYMDPALSIVLV 241

Qy   260 ILVLGTTLTILRDVILVLMEGTPKGVDFTAVRDLLL-SVEGVEALHSLHIWALTVAQPVL 318
          :||::||::|:||||||::|||||    :|:|||| :::||||:|||:|||:|::
Db   242 VLLHSVWPLLRESALILLQTVPTHIQVDAIQKRLLEKVDGVLAVHEEHVWQLAGDRIIA 301

Qy   319 SVHIAIAQNTDAQAVLKTPASSRLQGKFH---FHTVTTQIEDYSEDMKDCQACQGPS 371
          |:|   :::||:||:||:  ::||    |   ||:||:|:|
Db   302 SAHIRCRNLSEYMKI----AEKVKEFFHNEGIHSTTIQPE-FSE-IEGCNMSDGTS 351
```

RESULT 36
AAR04584
ID   AAR04584 standard; protein; 442 AA.
XX
AC   AAR04584;
XX
DT   19-SEP-1990 (first entry)
XX
DE   Protein product of plasmid p2N10 containing a novel yeast
DE   promoter.
XX
KW   p2N10; promotor; yeast; expression system; ds.
XX
OS   Saccharomyces cerevisiae.
XX
XX
PN   JP02092285-A.
XX
PD   03-APR-1990.
XX
PF   30-SEP-1988; 88JP-0246056.
XX
PR   30-SEP-1988; 88JP-0246056.
XX
PA   (MITU) MITSUBISHI KASEI CORP.
XX
DR   WPI; 1990-144899/19.
DR   N-PSDB; AAQ04338.
XX
PT   DNA fragment - shows good promotor activity in yeast.
XX
PS   Disclosure; ; ; Japanese.
XX
CC   Promoter is useful in various foreign gene expressing vectors.
XX
SQ   Sequence 442 AA;

Query Match   18.6%;   Score 358; DB 11; Length 442;
Best Local Similarity   25.2%;
Matches   97; Conservative   80; Mismatches 108; Indels 100; Gaps 13;

```
Qy    72 RQLYVASAICL--LFMIGEVVGGYLAHSLAVMTDAAHLLTDFASMLISLFSLWMS-SRPA 128
          :::|||   ::||||:||:::|:||||:||::::::::::||||::::::: :|||
Db     5 KELRIISLLTLDTVFFLLEITIIGYMSHSLALIADSFHMLNDIISLLVALWAVDVAKNRGP 64

Qy   129 TKTMNFGWQRAEILGALVSVLSIWVVTGVLVLVLAVERLISGDYEIDGGTMLITSGCAVAV 188
          :::|||||||||::||:  |:|||::|::|:|:|||||::::|:||:|| |::|::||||:
Db    65 DAKYTYGWKRAEILGALINAVFLIALCFCIMIEALQRLIE-PQEIQNPRLVLYGVGVAGLI 123

Qy   189 NIIMGLTL-HQSG----HGHSHGTTN------------------------- 209
          ||||||||  |:    ||:|:||:||
Db   124 SNVVGLFLFHDHGSDSVHSHGSVESGNNDLDIESNATHSHSHASLPNDNLAIDEDAIS 183

Qy   210 ------------------QQEENP-------------- 228
                            ::|||
Db   184 SPGPSGQIGEVLPQSVVNRLSNESQPLLNHDHDHSHESKKPGHRSLNMHGVFLHVLGDA 243

Qy   229 MQSMGVLVAAYILYFKPEYK---YVDPICTFVFSILVLGTTLTILRDVILVLMEGTPKGV 285
          ::|||:||||:||||||||     ||:||  |:||::||::|:|||||||::||||||:
Db   244 LGNIGVIAAALFIW-KTEYSWRYSDPIVSLLITIIIIFSSALPLSRRASRILLQATPSTI 302
```

Right column

```
Qy   286 DFTAVRDLLLSVEGVEALHSLHIWALTVAQPVLSVHIAIAQNTDAQAVLKTASSRLQGK- 344
          :::||: ::|||||||:|||:|||||:||::|:|||:||:::::|||:||:|||:  |
Db   303 SADQIQREILAVPGVIAVHDEHVWNITESIYIASIHVQI----DCAPDKEMSSAKLIRKI 358

Qy   345 FH---FHTVTIQIE-----DYSEDMK 362
          ||   |||:|||||     |:|:||:
Db   359 FHQHGIHSATVQPEFVSGDVNEDIR 383
```

RESULT 37
AAR95451
ID   AAR95451 standard; protein; 429 AA.
XX
AC   AAR95451;
XX
DE   Yeast OSR.
XX
DT   13-SEP-1996 (first entry)
XX
KW   OSR; S. cerevisiae strain DKD-5D-H; glutathione; peroxide resistance.
XX
OS   Saccharomyces cerevisiae strain DKD-5D-H.
XX
XX
PN   JP08070884-A.
XX
PD   19-MAR-1996.
XX
PF   02-SEP-1994; 94JP-0210205.
XX
PR   02-SEP-1994; 94JP-0210205.
XX
PA   (KYOW) KYOWA HAKKO KOGYO KK.
XX
DR   WPI; 1996-203162/21.
DR   N-PSDB; AAT26901.
XX
PT   Prodn. of glutathione using peroxidase resistant yeast - useful in
PT   the manufacture of drugs and foodstuffs
XX
PS   Disclosure; Page 6-8; 8pp; Japanese.
XX
CC   This sequence is encoded by the OSR gene from S. cerevisiae strain DKD-
CC   5D-H. The recombinant yeast containing the OSR DNA sequence was used
CC   in the preparation of glutathione by peroxide resistant microbes.
CC   Glutathione produced by the recombinants is secreted in to the
CC   culture medium and collected. It is useful in drugs and foodstuffs.
XX
SQ   Sequence 429 AA;

Query Match   18.6%;   Score 357; DB 17; Length 429;
Best Local Similarity   25.2%;
Matches   97; Conservative   80; Mismatches 108; Indels 100; Gaps 13;

```
Qy    72 RQLYVASAICL--LFMIGEVVGGYLAHSLAVMTDAAHLLTDFASMLISLFSLWMS-SRPA 128
          :::|||   ::||||:||:::|:||||:||::::::::::||||::::::: :|||
Db     5 KELRIISLLTLDTVFFLLEITIIGYMSHSLALIADSFHMLNDIISLLVALWAVDVAKNRGP 64

Qy   129 TKTMNFGWQRAEILGALVSVLSIWVVTGVLVLVLAVERLISGDYEIDGGTMLITSGCAVAV 188
          :::|||||||||::||:  |:|||::|::|:|:|||||::::|:||:|| |::|::||||:
Db    65 DAKYTYGWKRAEILGALINAVFLIALCFSIMIEALQRLIE-PQEIQNPRLVLYGVGVAGLI 123

Qy   189 NIIMGLTL-HQSG----HGHSHGTTN------------------------- 209
          ||||||||  |:    ||:|:||:||
Db   124 SNVVGLFLFHDHGSDSLHSHGSVESGNNDLDIESNATHSHSHASLPNDNLAIDEDAIS 183

Qy   210 ------------------QQEENP-------------- 228
                            ::|||
Db   184 SPGPSGQIGEVLPQSVVNRLSNESQPLLNHDHDHSHESKKPGHRSLNMHGVFLHVLGDA 243

Qy   229 MQSMGVLVAAYILYFKPEYK---YVDPICTFVFSILVLGTTLTILRDVILVLMEGTPKGV 285
          ::|||:||||:||||||||     ||:||  |:||::||::|:|||||||::||||||:
Db   244 LGNIGVIAAALFIW-KTEYSWRYSDPIVSLLITIIIIFSSALPLSRRASRILLQATPSTI 302
```

```
Qy  286 DFTAVRDLLLSVEGVEALHSLHIWALTVAQPVLSVSHIAIAQNTDAQAVLKTPASSRLQGK- 344
        :|:|| |||:||||| | |:|:|| || || |:|: |:|:|:| | |: |:|:|
Db  303 SADQIQREILAVPGVIAVHDFHVWNITESIYIASIHVQI----DCAPDKFMSSAKLIRKI 358

Qy  345 FH---FHTVTIQIE----DYSEDMK 362
        || ||:||||:| |:|||||
Db  359 FHQHGIHSATVQPEEVSGDVNEDIR 383
```

RESULT 38
AAG91156

ID AAG91156 standard; Protein; 318 AA.
XX
AC AAG91156;
XX
DT 26-SEP-2001 (first entry)
XX
DE C glutamicum protein fragment SEQ ID NO: 4910.
XX
KW Coryneform bacterium; amino acid synthesis; vitamin; saccharide;
KW organic acid synthesis.
XX
OS Corynebacterium glutamicum.
XX
PN EP1108790-A2.
PD 20-JUN-2001.
XX
PF 18-DEC-2000; 2000EP-0127688.
XX
PR 16-DEC-1999;   99JP-0377484.
PR 07-APR-2000; 2000JP-0159162.
PR 03-AUG-2000; 2000JP-0280988.
XX
PA (KYOW) KYOWA HAKKO KOGYO KK.
XX
PI Nakagawa S, Mizoguchi H, Ando S, Hayashi M, Ochiai K, Yokoi H;
PI Tateishi N, Senoh A, Ikeda M, Ozaki A;
XX
DR WPI; 2001-376931/40.
DR N-PSDB; AAH66375.
XX
PT Novel polynucleotides derived from Coryneform bacteria, for identifying
PT mutation point of a gene, measuring expression of a gene, analysing
PT expression profile or pattern of a gene and identifying homologous gene
PT -
XX
PS Claim 17; SEQ ID NO: 4910; 246pp + Sequence Listing; English.
XX
CC The present invention provides a number of nucleotide and protein
CC sequences from the Coryneform bacterium Corynebacterium glutamicum. These
CC are useful for identifying the mutation point of a gene derived from a
CC mutant of coryneform bacterium, measuring expression amount and
CC analysing the expression profile or identifying a homologue of a gene derived
CC from Coryneform bacterium, and identifying a homologue of a gene derived
CC from coryneform bacterium. Coryneform bacteria are useful for producing
CC amino acids, nucleic acids, vitamins, saccharides and organic acids,
CC particularly L-lysine. The present sequence is a protein described
CC in the exemplification of the invention.
CC Note: The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from the
CC European Patent Office.
XX
SQ Sequence 318 AA;

```
Qy  39 DLQAIELAAQSNHHCHAQKGPDSHCDPKKGKAQORLVASAICLLFMIGEVVGGYLAHSL 98
       ||| |||: |:  | |:|  ||
Db  5  DLQQRSYAHNPDGHDHSHDGLGHSHAPSSLKA---LFAVIIFTSIIFLAELIAGLISGSL 61
```

Query Match 18.2%; Score 350; DB 22; Length 318;
Best Local Similarity 27.5%; Pred. No. 7.1e-29;
Matches 92; Conservative 72; Mismatches 141; Indels 30; Gaps 9;

```
Qy   99 AVWTDAAHLLTDFASMLTSLFSLWMSSRPATKTMNFGWQRAELIGALVSVLSIWVVTGVL 158
         :||| ||||: |:|| ||| ||  ||  |: |:|:||  ||::| |:|:|:|:  |::|:
Db   62 ALLADAMHMLSDSTGLIIAAVAMLIGRRARTSRATYGYKRAEVLAAMVNAT---VVTALS 118

Qy  159 VYLAVERL--ISGDYEIDGGTMLITSCGAVAVNIIMGLTL--HOSGHGHSHGTTNQQEEN 214
         | ||  |  :|| :|: :||:|:| |  ||:| ||:  |   |  |   |
Db  119 WIVVEAIMRLGKDLEIQTNLMLIVAVIGFVTNGISALVLMRHODG-------------N 165

Qy  215 PSVRAAFIHVIGDFMQSMGVLVAAYLIYFKPEYKYVDPICTFVFSILVLGTTLTILRDVI 274
         :|:|||:|||: || :||||:||| | ||   ||| |:|:|: |:|: ||:||:| : :
Db  166 INMRGAFLHVLSDMLGSVAVIIAGLVIRY-TGWMPADTIASIAIAAIIIPRAFSLLKEAL 224

Qy  275 LVLMEGTPKGVDETAVRDLLLSVEGVEALHSLHIWALTVAQPVLSVSHIAIAQNTD--AQ 331
         :|:| |  :| :  ||: |: |  | |||||||||| : :|:: |   |:  | |   :|
Db  225 NILLERVPTGAEPAEVDAALRKVPGVSDVHDLHIWSIDGKEILATVHLVVDSSTNQLHSC 284

Qy  332 AVLKTASSRLQGKFHFHTVTIQIE--DYSEDMKDC 364
         ||| ||: ::| :  |:|::||:|  |: :|: :|
Db  285 GVLDRAEAELSKLGILHS-TIQLESADHSDHESVC 318
```

RESULT 39
AAB76797

ID AAB76797 standard; Protein; 318 AA.
XX
AC AAB76797;
XX
DT 11-APR-2001 (first entry)
XX
DE Corynebacterium glutamicum MCT protein SEQ ID NO:576.
XX
KW Corynebacterium glutamicum; brevibacterium lactofermentum; MCT;
KW membrane construction and membrane transport protein; petroleum spill;
KW hydrocarbon degradation; gram positive aerobic bacterium; marker;
KW identification; microorganism; fine chemical production; transformation;
KW genome mapping; genetic engineering.
XX
OS Corynebacterium glutamicum.
XX
PN WO200100805-A2.
PD 04-JAN-2001.
XX
PF 23-JUN-2000; 2000WO-IB00926.
XX
PR 25-JUN-1999;   99US-0141031.
PR 08-JUL-1999;   99DE-1031454.
PR 08-JUL-1999;   99DE-1031478.
PR 08-JUL-1999;   99DE-1031563.
PR 09-JUL-1999;   99DE-1032122.
PR 09-JUL-1999;   99DE-1032124.
PR 09-JUL-1999;   99DE-1032125.
PR 09-JUL-1999;   99DE-1032128.
PR 09-JUL-1999;   99DE-1032180.
PR 09-JUL-1999;   99DE-1032182.
PR 09-JUL-1999;   99DE-1032190.
PR 09-JUL-1999;   99DE-1032191.
PR 09-JUL-1999;   99DE-1032209.
PR 09-JUL-1999;   99DE-1032212.
PR 09-JUL-1999;   99DE-1032227.
PR 09-JUL-1999;   99DE-1032228.
PR 09-JUL-1999;   99DE-1032229.
PR 09-JUL-1999;   99DE-1032230.
PR 14-JUL-1999;   99DE-1032927.
PR 14-JUL-1999;   99DE-1033005.
PR 14-JUL-1999;   99DE-1033006.
PR 27-AUG-1999;   99DE-1040764.
PR 27-AUG-1999;   99DE-1040765.
PR 27-AUG-1999;   99DE-1040766.
PR 27-AUG-1999;   99DE-1040830.
PR 27-AUG-1999;   99DE-1040831.
PR 27-AUG-1999;   99DE-1040832.

PR   27-AUG-1999;   99DE-1040833.
PR   31-AUG-1999;   99DE-1041378.
PR   31-AUG-1999;   99DE-1041379.
PR   31-AUG-1999;   99DE-1041395.
PR   03-SEP-1999;   99DE-1042077.
PR   03-SEP-1999;   99DE-1042078.
PR   03-SEP-1999;   99DE-1042079.
PR   03-SEP-1999;   99DE-1042088.
XX
PA   (BADI) BASF AG.
XX
PI   Pompejus M, Kroeger B, Schroeder H, Zelder O, Haberhauer G;
XX
PR   WPI; 2001-071486/08.
DR   N-PSDB; AAF68030.
XX
PT   Corynebacterium glutamicum nucleic acids encoding membrane construction
PT   and membrane transport proteins or their portions, useful for typing or
PT   identifying C. glutamicum or related bacteria, and as markers for
PT   transformation -
XX
PS   Claim 20; Page 964-965; 1111pp; English.
XX
CC   AAF67743 to AAF68080 encode the Corynebacterium glutamicum membrane
CC   construction and membrane transport (MCT) proteins given in AAB76510 to
CC   AAB76847. The MCT nucleic acids and proteins are useful in the
CC   identification of microorganisms which can be used to produce fine
CC   chemicals, for modulating fine chemical production in C. glutamicum or
CC   related bacteria (e.g. Brevibacterium lactofermentum), the typing or
CC   identification of C. glutamicum or related bacteria, as reference points
CC   for mapping C. glutamicum genome, and as markers for transformation.
CC   AAF68082 and AAF68082 represent sequencing primers which are used in an
CC   example from the present invention.
XX
SQ   Sequence 318 AA;

Query Match   18.2%; Score 350; DB 22; Length 318;
Best Local Similarity   27.5%;
Matches   92; Conservative 72; Mismatches 141; Indels 30; Gaps   9;

Qy   39 DLQAIELAAQSNHHCHAQKGPDSHCDPKKGKAQROLYVASAICLLFMIGEVVGGYLAHSL   98
          ||| :|||:   :|::| :||| :|::|: |:|::||||| ||
Db   5 DLQQRSYAHNPDGHDHSHDGLGHSHAPSSLKA---LFAVIIFTSIIFLAELIAGLISGSL 61

Qy   99 AVWTDAAHLLTDFASMLISLFSLWMSSRPATKTMNFGWQRAEILGALVSVLSIWVVTGVL 158
          ||| || || |:||| :: :: |:|| | ||:|||||||||||||: |||:|||
Db   62 ALLADAMHMLSDSTGLIIAAVAMLIGRRARTSRATYGYKRAEVLAAMVNAT---VVTALS 118

Qy   159 VYLAVERL--ISGDYEIDGGTMLITSGCAVAVANIIMGLTL--HOSGHGHSHGTTNQEEN 214
          |:||||| |:|| |||||||||||||||| |||||||| ||:||||||:
Db   119 VWIVVEAIMRLGKDLEIQTNLMLIVAVIGFVTNGISALVLMRHODG---------N 165

Qy   215 PSVRAAFIHVIGDFMQSMCVLVAAAYILYFKPEYKYVDPICTFVFSILVLGTTLTTLRDVI 274
          |:|:|||||||||||||:||||||| :||| ||: || ::||:: |||:|||||| :||:
Db   166 INMRGAFLHVLSDMLGSVAVITAGLVIRY-TGWMPADTIASIAIAAIIPRAEFSLLKEAL 224

Qy   275 LVLMEGTPKGVDFTAVRDLLSVEGVEALHSLHIWALTVAQPVLSVHIAIAQNTD---AQ 331
          :|:| ||: || |:|:: |:|||||:||||| |||||||||||:|::: || ||
Db   225 NILLERVPFGAEPAEVDAALRKVPGVSDVHDLHIWSIDGKEILATVHLVVDSSTNQLHSC 284

Qy   332 AVLKTASSRLQGKFHFHVTIQIE--DYSEDMKDC 364
          ||||||||||: |:|| |||| | ||| |||:
Db   285 GVLDRAEALSKLGILHS-TIQLESADHSDHESVC 318

SQ   Sequence 318 AA;
XX
DT   26-NOV-2001 (first entry)

DE   Amino acid sequence of a human transporter protein.
XX
KW   Human; transporter protein; zinc transporter; vesicular transporter.
XX
OS   Homo sapiens.
XX
PN   WO20016A878-A2.
XX
PD   07-SEP-2001.
XX
PF   28-FEB-2001; 2001WO-US06462.
XX
PR   29-FEB-2000; 2000US-0185956.
XX
PA   (LEXI-) LEXICON GENETICS INC.
XX
PI   Walke DW, Wilganowski NL, Hu Y, Kieke JA, Zambrowicz B, Revelli J;
PI   Turner CA;
XX
DR   WPI; 2001-570693/64.
DR   N-PSDB; AAH78269.
XX
PT   Novel polynucleotides useful for detection of mutant proteins or
PT   inappropriately expressed proteins for the diagnosis of disease, drug
PT   screening, and therapeutics -
XX
PS   Disclosure; Page 47-48; 49pp; English.
XX
CC   The present sequence represents a human transporter protein. The protein
CC   is similar to zinc transporter and vesicular transporter proteins. The
CC   transporter proteins and their polynucleotides, as well as antibodies
CC   specific for the proteins may be used for the detection of mutant
CC   proteins or inappropriately expressed proteins for the diagnosis of
CC   disease. In addition they may be used to screen for drugs and as
CC   therapeutics.
XX
SQ   Sequence 507 AA;

Query Match   18.1%; Score 347.5; DB 22; Length 507;
Best Local Similarity   24.6%;
Matches 108; Conservative 71; Mismatches 117; Indels 143; Gaps 13;

Qy   68 GKAQROLYVASAICLLFMIGEVVGGYLAHSLAVMTDAAHLLTDFASMLISLFSLWMSSRP 127
          |::| | |: || ||:||:||:||| || || ::||| :||:||||||||||||:
Db   5 GRNRGRLLCMLALTFNFMVLEVVVSRVTSSLAMLSDSFHMLSDVLALVVALVAERFARRT 64

Qy   128 -ATKTMNFGWQRAEILGALVSVLSIWVVTGVLVYLAVERLISGDYEIDGGTMLITSGCAV 186
           || |||||||||||||||:|||||||||:|:|:|| |||| |:|||:||||||||
Db   65 HATQKNTFGWIRAEVMGALVNAIFLTGLCFAILLEAIERFIE-PHEMQQPLVVLGVGVAG 123

Qy   187 AVNIIMGLTL------HQSGHGHSHG------------ 206
          |::|||||: ||:|||||
Db   124 LLVNVLGLCLFHHHSGFSQDSGHGHSHGGHGHGHGLPKGPRVKSTRPGCSSDINVAPGEOQ 183

Qy   207 -------------TTNQ--------QEENP------- 215
                       ::| |::|
Db   184 PDQEEFNTLVANTSNSNGLKLDPADPENPRSGDTVEVQVQVGNLVREPDHMELEEDRAGQL 243

Qy   216 SVRAAFIHVIGDFMQSMGVLVAAAYILYFK-----------------PE---------- 246
          |:|||||||||||||:|| ||||| :||| ||
Db   244 NMRGVFLHVLGDALGSVIVVVNALVFYFSWKGCSEGDFCVNPCFPDPCKAFVEIINSTHA 303

Qy   247 --YK------YVDPICTFVFSILVLGTTLTTLRDVILVLMEGTPRGVDFTAVRDLLLSV 303
            :| :|| |:||:: |||:||||| :||:|::| ||: || |:|:: |:|
Db   304 SLYEAGPCWLYLDPTLCVWWVCILLIYTTVPLLKESALILLQTVPKQIDIRNLIKELRNV 363

Qy   355 EDYSEDMKD----CQ-ACQ 368
          |||| ||| || :||
Db   364 EGVEEVHELHVWQLAGSRIIATAHICEDPYSYMEVAKT----IKDVFHNGIHATTIQP 419

Qy |: |||:
Db   420 EFASVGSKSSVVPCELACR 438

XX
DT   26-NOV-2001 (first entry)

Search completed: March 18, 2003, 15:40:43
Job time : 47 secs